

UNIVERSIDADE FEDERAL DO PARANÁ

DIEGO MACIEL GERONIMO

APLICAÇÃO DE FERRAMENTAS QUIMIOMÉTRICAS NA INDÚSTRIA
COSMÉTICA: UM ESTUDO SOBRE ANÁLISE DE PRODUTOS, CONTROLE DE
QUALIDADE E DETERMINAÇÃO MULTIVARIADA DE COMPOSTOS

CURITIBA
2016

DIEGO MACIEL GERONIMO

APLICAÇÃO DE FERRAMENTAS QUIMIOMÉTRICAS NA INDÚSTRIA
COSMÉTICA: UM ESTUDO SOBRE ANÁLISE DE PRODUTOS, CONTROLE DE
QUALIDADE E DETERMINAÇÃO MULTIVARIADA DE COMPOSTOS

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Química, no Curso de Pós-Graduação em Química, Setor de Ciências Exatas, Universidade Federal do Paraná.

Área de concentração: Química Analítica.

Orientadora: Profa. Dra. Noemi Nagata
Coorientador: Prof.Dr. Patricio Peralta-Zamora

CURITIBA
2016

Catálogo na Fonte: Sistema de Bibliotecas, UFPR
Biblioteca de Ciência e Tecnologia

G377a Geronimo, Diego Maciel
Aplicação de ferramentas quimiométricas na indústria cosmética:
um estudo sobre análise de produtos, controle de qualidade e
determinação multivariada de compostos [Recurso eletrônico] / Diego
Maciel Geronimo – Curitiba, 2016.

Dissertação (mestrado) - Universidade Federal do Paraná, Setor de
Ciências Exatas Programa de Pós-graduação em Química, 2016.

Orientadora: Noemi Nagata
Coorientador: Patricio Peralta-Zamora

1. Cosméticos. 2. Quimiometria. I. Universidade Federal do Paraná.
II. Nagata, Noemi. III. Peralta-Zamora, Patricio. IV. Título.

CDD: 668.55

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

TERMO DE APROVAÇÃO

**APLICAÇÃO DE FERRAMENTAS QUIMIOMÉTRICAS NA INDÚSTRIA
COSMÉTICA: UM ESTUDO SOBRE ANÁLISE DE PRODUTOS,
CONTROLE DE QUALIDADE E DETERMINAÇÃO MULTIVARIADA DE
COMPOSTOS**

por

DIEGO MACIEL GERONIMO

Dissertação aprovada como requisito parcial para obtenção do grau de
Mestre no Programa de Pós-Graduação em Química,
pela Comissão Examinadora composta por:

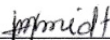
Orientador:



Prof.ª Dr.ª Noemi Nagata
Dep. de Química – UFPR



Prof. Dr. Marco Flôres Ferrão
Instituto de Química – UFRGS



Prof.ª Dr.ª Iara Messerschmidt
Dep. de Química – UFPR

Curitiba, 25 de fevereiro de 2016.

**Dedico este trabalho a meus pais Jamil e
Marli, exemplos de superação e
persistência.**

AGRADECIMENTOS

A todos que de alguma maneira contribuíram para a realização deste trabalho, em especial a minha orientadora, Professora Dra. Noemi Nagata, pela paciência, apoio, carinho, amizade e por acreditar em mim e neste projeto. Por contribuir para o meu crescimento pessoal e profissional e por ser um exemplo de competência e dedicação.

Ao meu coorientador, Professor Dr. Patricio Guillermo Peralta Zamora, pelo suporte e contribuições para este trabalho. Pelos palpites, recomendações e por sempre me desafiar a pensar fora da hipercaixa.

Ao Professor Dr. Bruno José Gonçalves da Silva, pela atenção e disposição constante em ajudar e por suas valiosas contribuições desde o projeto inicial, relatório e exame de qualificação, que contribuíram muito para a evolução deste documento.

À Professora Dra. Iara Messerschmidt por ter aceitado prontamente o convite para participar da banca de defesa de mestrado e por contribuir com sua ampla experiência e conhecimento, trazendo sugestões de melhoria que ajudaram este trabalho ser o que acabou se tornando.

Ao Professor Dr. Marco Flores Ferrão por ter aceitado participar da banca de defesa de mestrado e pelas contribuições para o trabalho, fornecendo excelentes sugestões de melhoria e apontando novos e interessantes caminhos.

Aos meus pais, Jamil e Marli, por todo o seu esforço durante as suas vidas dedicadas a fornecer aos seus filhos condições de conforto e educação para que pudéssemos perseguir nossos sonhos.

Ao meu irmão Bruno, pelo exemplo de dedicação e perseverança.

À Luiza, pela compreensão, carinho e companheirismo em todos os momentos. Pelo tempo dedicado a mim, mesmo quando faculdade, trabalho, voluntariado, família, amigos e todo um mundo de ocupações concorriam pelos minutos dos seus dias. Pelo sorriso e abraço que me deram força nos momentos mais difíceis.

Aos amigos das famílias GAQQ e TECNOTRATER e agregados pelo apoio com os experimentos e manutenção da saúde mental e emocional nos momentos de correria: Alecsandra Santos, Arlene Nossol, Bianca Amaral, Daniele Scheres Firak, Ellery Regina Garbelini, Graziela Da Silva Costa, Ivan de Barros, Jéssica

Schneider, Larissa Bach, Priscila Lagner, Rayta Paim Horta, Rodrigo Barbosa Pinto, Sandra Stets e Suzamar Moura Costa Rosa. Com agradecimento especial às amigas Bianca, Ale e Sandra pelo cuidado e paciência com minhas infinitas dúvidas.

Ao Grupo Boticário, cuja visão e jeito inovador me abriram portas e sempre me ajudaram a atingir meus objetivos pessoais e profissionais; e cuja estrutura tornou possível boa parte do presente trabalho.

A todos os colegas de empresa pelo companheirismo e auxílio no desenvolvimento do trabalho e discussões correlatas. Em especial meus gestores e mentores do presente e do passado Marina Adami, Gustavo Dieamant, Carlos Praes, Alexandre Silva e Rodrigo Madalosso pela mentoria, flexibilidade e compreensão. A Sandra, Carlos, Luna e demais colegas de Desenvolvimento Analítico pelos ensinamentos e parceria nas análises. A Ana Paula, Adilson, Camila, Monique, Isabela e demais amigos que participaram mais diretamente dos trabalhos e que se dispuseram a me auxiliar nas diversas frentes deste projeto.

Aos amigos de ICT: Aline, Amanda, Bruna, Felipe B, Felipe K, Frederico, Laís, Mariana, Otto, Vick e Valentine. Em especial ao Fred pelas aulas de R e de princípios como o de duas latas.

À equipe da coordenação da pós-graduação: Professora Dra. Jaísa Fernandes Soares, Professor Dr. Fábio Souza Nunes, Marcelino Câmara e Fernanda Gabardo Dias Pinheiro pela eficiência, paciência e atenção.

Às minhas amigas da vida inteira Bianca, Karina, Laiane, Samantha e Thatiane, por estarem sempre presentes, mesmo à distância.

Aos amigos que fiz durante a graduação – e que espero que continuem sendo meus amigos para sempre – Aline, Barbara, Camila, Daniela, Dásio, Douglas, João, Letícia, Lucas, Maurício, Talita, Vanessa, Valmir e demais colegas.

Aos amigos de todos os tempos do LECOSIN e agregados: Angelita, Laieli, Marina, Paula, Celso, Gustavo, Kahlil, Odilon, Ricardo e Vinícius.

A todos que fizeram parte deste trabalho, mas que por fraqueza de memória tenha esquecido de nomear.

“Essencialmente, todos os modelos estão
errados, mas alguns são úteis”.

George Edward Pelham Box

RESUMO

A indústria cosmética é uma das que mais cresce em todo o mundo em termos de faturamento e volume de vendas, sendo protagonista na pesquisa, desenvolvimento e inovação em vários campos da ciência, como nanotecnologia e segurança. Em razão das numerosas necessidades deste segmento industrial e da usual disponibilidade de ferramentas instrumentais de análise, o uso de técnicas quimiométricas pode ser de grande valor para o desenvolvimento de rotinas de controle de qualidade, agilizando as respostas e, conseqüentemente, a tomada de decisão. Nesse contexto, algumas aplicações das ferramentas quimiométricas foram realizadas. Para verificar a massa base empregada na confecção de sabonetes em barra (presença de massa base animal, vegetal ou sintética), foi empregada a técnica de Espectroscopia de Infravermelho por Reflectância Total Atenuada (ATR-FTIR) associada aos métodos de calibração multivariada como a Análise Discriminatória por Regressão de Mínimos Quadrados Parciais (PLS-DA) e Modelagem Independente por Analogia de Classe (SIMCA), os quais levaram a 100 % de acerto na classificação de 9 amostras de sabonete que não participaram da construção do modelo, evidenciando a validade dos métodos propostos para diferenciar amostras com alto grau de similaridade. Na PLS-DA, o modelo multivariado foi construído com a faixa espectral entre 1850 e 800 cm^{-1} , dados centrados na média e utilizando três variáveis latentes, as quais explicam 99,15 % da informação contida na matriz X e 92,15 % da informação contida na matriz Y. O modelo SIMCA foi construído com os mesmos pré-processamentos do PLS-DA e empregando-se quatro componentes principais para as classes sintética e vegetal e 3 para a classe animal. Adicionalmente, foi desenvolvida uma metodologia de triagem de testes microbiológicos visando a redução nos tempos de desenvolvimento de produtos cosméticos. A metodologia construída apontou que o teste de Halo de Inibição para o fungo *Aspergillus niger* constitui um marcador preditivo adequado para prever os resultados de um sistema conservante frente ao *Challenge Test*, com valores de halo entre 11 e 18 mm apontando que um resultado eficaz pode ser esperado e reduzindo de 28 para 2 dias o período de comparação de diferentes conservantes para uso em produtos. Outra frente de trabalho incluiu o desenvolvimento de uma metodologia quantitativa via PLS do teor do conservante Caprilato de Poliglicerila em amostras de creme hidratante (dados centrados na média e utilizando duas variáveis latentes, as quais explicam 96,69 % da informação contida na matriz Y e 97,24 % da informação contida na matriz X, levando a um valor de $R^2 = 0,968$ e a um erro máximo de 7,92% para a concentração do conservante em amostras externas ao conjunto de calibração). Por fim, foi elaborado um sistema baseado nos métodos de HCA e PCA (com dados autoescalados), o qual foi empregado para agrupar 38 emolientes de acordo com as características físico-químicas mais relevantes, sendo capaz de recomendar os melhores substitutos para cada amostra testada.

Palavras-chave: Cosméticos; Quimiometria; Análises.

ABSTRACT

The cosmetic industry is one of the fastest growing markets in the world in terms of revenue and sales volume, being protagonist in research, development and innovation in various fields of science such as nanotechnology and safety assessment. Due to the many needs of this industrial segment and the usual availability of instrumental analysis tools, using chemometric techniques can be of great value to the development of quality control routines, speeding up the answers and, consequently, the decision-making process. In this context, applications of chemometric tools were used. To verify the origin of bar soap raw materials (animal, vegetable or synthetic), it was applied the Infrared Spectroscopy technique with Attenuated Total Reflectance (ATR-FTIR) in association with the multivariate methods of Partial Least Squares Discriminatory Analysis (PLS-DA) and Soft Independent Modeling of Class Analogy (SIMCA), which led to a 0 % error in the classification of 9 bar soap samples that weren't included in the model construction. In the regression via PLS-DA, the multivariate model was built with the spectral range between 1850 and 800 cm^{-1} , mean centered data and using three Latent Variables, which explain 99.15 % of the Information contained in the X matrix and 92.15 % of the information contained in the Y matrix. The SIMCA model was built using the same pre-processing applied to the PLS-DA model and employing four Principal Components for the synthetic and vegetable classes and 3 for the animal class. In addition, a microbiological screening test method aimed at reducing cosmetic products development time has been developed. The methodology built pointed out that the inhibition zone test for the *Aspergillus niger* fungus is a suitable predictive marker for the outcome in the Challenge Test of a preservative system, with halo values between 11 and 18 mm pointing out that effective results can be expected, reducing from 28 to 2 days the preservative systems comparison period. Other work front included the development of a quantitative methodology via PLS aiming to determine the preservative (Polyglyceryl Caprylate) content in samples of a moisturizing cream (mean centered data and using two latent variables, which explain 96.69 % of the information contained in the Y array and 97.24 % of the information contained in the X array, leading to a value of $R^2 = 0.968$ and a maximum error of 7.92 % for the preservative concentration in samples external to set calibration). Finally, it was designed a system based on HCA and PCA methods (with autoscaled data), which was used to group 38 emollients according to their most relevant physical and chemical characteristics, being able to recommend the best substitutes for each sample tested.

Key-words: Cosmetics; Chemometrics; Analysis.

LISTA DE FIGURAS

FIGURA 1 – EXEMPLO DE REPRESENTAÇÃO DE AGRUPAMENTOS VIA DENDROGRAMA.....	25
FIGURA 2 - ILUSTRAÇÃO DO PROCESSO DE REDUÇÃO DE DIMENSIONALIDADE. a) CONJUNTO TRIDIMENSIONAL DE DADOS; b) PROPOSTA DE 3 COMPONENTES PRINCIPAIS; E c) PROJEÇÃO DOS ELEMENTOS ORIGINAIS EM FUNÇÃO DE DUAS PCS (KAVRAKI, 2007)	26
FIGURA 3 - CONJUNTO DE DADOS BIDIMENSIONAIS MOSTRANDO OS EIXOS DAS VARIÁVEIS ORIGINAIS (λ_1 e λ_2), DAS COMPONENTES PRINCIPAIS (PC1 E PC2) E A REPRESENTAÇÃO DE SCORES (T) E LOADINGS (P) (FERREIRA <i>et al.</i> , 1999)	27
FIGURA 4 – (a) REPRESENTAÇÃO GRÁFICA DOS PARÂMETROS T^2 DE HOTELLING E Q RESIDUAL PARA TRÊS VARIÁVEIS E DUAS PCs. (b) MODELO SIMCA PARA TRÊS CLASSES, SENDO QUE CADA CLASSE APRESENTA UM NÚMERO DIFERENTE DE COMPONENTES PRINCIPAIS E AMOSTRAS DE CALIBRAÇÃO. AS DISTÂNCIAS DO PONTO k PARA CADA UMA DAS CLASSES TAMBÉM ESTÁ REPRESENTADA (HALL; KENNY, 2007).....	33
FIGURA 5 – VARIÂNCIA ACUMULADA COM O AUMENTO DO NÚMERO DE COMPONENTES PRINCIPAIS	42
FIGURA 6 – APRESENTAÇÃO GRÁFICAS DAS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS. AMOSTRAS EFICIENTES E INEFICIENTES NO CHALLENGE TEST	42
FIGURA 7 - LOADINGS DAS PCS 1 E 2	43
FIGURA 8 – RESULTADOS DA ÁRVORE DE DECISÃO.....	44
FIGURA 9 –INFORMAÇÕES SOBRE O CONSERVANTE CAPRILATO DE POLIGLICERILA	47
FIGURA 10 – (A) ESPECTROS DE INFRAVERMELHO DAS 11 AMOSTRAS DO CONJUNTO DE CALIBRAÇÃO. (B) COMPARAÇÃO ENTRE O ESPECTRO MÉDIO DAS AMOSTRAS E O ESPECTRO DO PADRÃO DO CONSERVANTE	51
FIGURA 11 – RMSECV VS. NÚMERO DE VARIÁVEIS LATENTES PARA O MODELO.....	53
FIGURA 12 – VETOR DE REGRESSÃO DO MODELO DE PLS PARA QUANTIFICAÇÃO DO CAPRILATO DE POLIGLICERILA.....	53
FIGURA 13 – REPRESENTAÇÃO GRÁFICA DA ANÁLISE DE RESÍDUOS STUDENTIZADOS <i>VERSUS LEVERAGE</i>	54

FIGURA 14 - COMPARAÇÃO ENTRE AS OS VALORES DE CONCENTRAÇÃO MEDIDOS CONTRA OS PREVISTOS PARA AS AMOSTRAS DE CREME. VALOR DE $R^2=0,968$	55
FIGURA 15 – DENDROGRAMA OBTIDO PARA AS 38 AMOSTRAS DE EMOLIENTES. A ESCALA ABAIXO DO DENDROGRAMA INDICA OS ÍNDICES DE DISTÂNCIA ENTRE AS AMOSTRAS (SIMILARIDADE MÁXIMA PARA DISTÂNCIA IGUAL A ZERO).....	64
FIGURA 16 – APRESENTAÇÃO GRÁFICA DOS SCORES (NÚMEROS EM CORES) E <i>LOADINGS</i> (FLECHAS EM VERMELHO) PARA AS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS	65
FIGURA 17 – ESPECTROS DE FTIR REPRESENTATIVOS DOS DIFERENTES GRUPOS DE SABONETES	74
FIGURA 18 - REGIÃO DESTACADA DO ESPECTRO DE FTIR DAS AMOSTRAS	75
FIGURA 19 – APRESENTAÇÃO GRÁFICA DAS TRÊS PRIMEIRAS COMPONENTES PRINCIPAIS. AMOSTRAS PROVENIENTES DE SABONETES CONTENDO MATÉRIA PRIMA (A) ANIMAL; (V) VEGETAL; E (S) SINTÉTICA..	76
FIGURA 20 - <i>LOADINGS</i> DAS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS (PCS 1 E 2)	77
FIGURA 21 – RMSECV VS. NÚMERO DE VARIÁVEIS LATENTES (a). VARIÂNCIA CAPTURADA POR CADA VARIÁVEL LATENTE (b).....	79
FIGURA 22 – VETOR DE COEFICIENTES DE REGRESSÃO DO MODELO DE PLS-DA PARA CLASSIFICAÇÃO DA ORIGEM DA MATÉRIA PRIMA EMPREGADA EM SABONETES EM BARRA.....	80
FIGURA 23 – REPRESENTAÇÃO GRÁFICA DA ANÁLISE DE RESÍDUOS STUDENTIZADOS <i>VERSUS LEVERAGE</i>	81
FIGURA 24 - COMPARAÇÃO ENTRE AS CLASSIFICAÇÕES REAIS	82
FIGURA 25 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 1 - MASSA SINTÉTICA (a). Q VS T^2 PARA QUATRO COMPONENTES PRINCIPAIS DA CLASSE 1. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 1 E EM VERMELHO AS AMOSTRAS DAS CLASSES 2 E 3 (b)	84
FIGURA 26 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 2 – MASSA VEGETAL (a). Q VS T^2 PARA QUATRO COMPONENTES PRINCIPAIS DA CLASSE 2. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 2 E EM VERMELHO AS AMOSTRAS DAS CLASSES 1 E 3 (b)	85
FIGURA 27 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 3 – MASSA ANIMAL (a). Q VS T^2 PARA TRÊS COMPONENTES PRINCIPAIS DA CLASSE 3. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 3 E EM VERMELHO AS AMOSTRAS DAS CLASSES 1 E 2 (b)	86

LISTA DE TABELAS

TABELA 1 - REGIÕES ESPECTRAIS NO INFRAVERMELHO	22
TABELA 2 – SISTEMAS CONSERVANTES ESTUDADOS.....	38
TABELA 3 – RESULTADOS DOS TESTES MICROBIOLÓGICOS PARA OS SISTEMAS CONSERVANTES.....	40
TABELA 4 - COMPOSIÇÃO QUALITATIVA DO CREME HIDRATANTE EMPREGADO.....	49
TABELA 5 - CONCENTRAÇÃO DE CAPRILATO DE POLIGLICERILA NAS AMOSTRAS.....	50
TABELA 6 - VALORES REAIS E PREVISTOS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA.....	55
TABELA 7 - CLASSIFICAÇÃO QUÍMICA DOS EMOLIENTES SELECIONADOS.....	60
TABELA 8 - ANÁLISES REALIZADAS PARA A CARACTERIZAÇÃO DOS EMOLIENTES	61
TABELA 9 – RESULTADOS DAS ANÁLISES DOS EMOLIENTES.....	63
TABELA 10 – RELAÇÕES OBTIDAS	67
TABELA 11 - CARACTERÍSTICAS DOS ÓLEOS E GORDURAS GERALMENTE EMPREGADOS EM SABÕES.....	70
TABELA 12 - PRINCIPAIS DIFERENÇAS COMPOSICIONAIS ENTRE OS ÓLEOS E GORDURAS MAIS EMPREGADOS NA FABRICAÇÃO DE SABONETES	71
TABELA 13 - VALORES MEDIDOS E PREVISTOS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA.....	83
TABELA 14 - CLASSES REAIS E PREVISTAS PARA O CONJUNTO DE VALIDAÇÃO INTERNA.....	87
TABELA 15 - CLASSES REAIS E PREVISTAS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA.....	87

LISTA DE SIGLAS E ABREVIATURAS

ABIHPEC - Associação Brasileira da Indústria de Higiene Pessoal Perfumaria Cosméticos

ANVISA - Agência Nacional de Vigilância Sanitária

ATR - Reflectância Total Atenuada

BHT - Hidroxitolueno butilado

cP - Centipoise (unidade de viscosidade)

CV - Validação Cruzada (Cross Validation)

D - Distância

EDTA - Ácido Etilenodiamino Tetra-Acético

FAR - Infravermelho distante

FTIR - Espectroscopia no infravermelho com transformada de Fourier

HPLC - Cromatografia em fase Líquida de Alta Eficiência

HCA - Análise por Agrupamentos Hierárquicos

INCI - Nomenclatura Internacional de Ingredientes Cosméticos

m - Massa

MID - Infravermelho médio

NIPALS - Mínimos Quadrados Parciais Não Iterativos

NIR - Infravermelho Próximo

PC - Componente Principal

PCA - Análise por Componentes Principais

PCR - Regressão por Componentes Principais

pH - Potencial Hidrogeniônico

PLS - Mínimos Quadrados Parciais

PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais

PLSR Regressão por Mínimos Quadrados Parciais

PRESS - Soma dos Quadrados dos Erros de Previsão

Q - Distância ortogonal entre uma amostra e o plano definido pelas componentes principais

R^2 - Coeficiente de determinação

RMSEC - Raiz Quadrada do Erro Médio Quadrático da Calibração

RMSECV - Raiz Quadrada do Erro Médio Quadrático da Validação Cruzada

RMSEP - Raiz Quadrada do Erro Médio Quadrático da Previsão

S - Desvio Padrão

SIMCA - Modelagem Independente por Analogia de Classe

T^2 - Resíduos (Hotelling)

UV-Vis - Região do Ultravioleta-Visível

LV - Variável Latente

Vs - Versus

SÚMARIO

1. INTRODUÇÃO	18
2. REVISÃO DA LITERATURA.....	19
2.1. COSMÉTICOS: HISTÓRIA E PANORAMA ECONÔMICO	19
2.2. CONTROLE DE QUALIDADE NA INDÚSTRIA COSMÉTICA.....	20
2.3. ESPECTROSCOPIA NO INFRAVERMELHO	21
2.4. QUIMIOMETRIA	23
2.5. ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)	24
2.6. ANÁLISE POR COMPONENTES PRINCIPAIS – PCA	25
2.7. REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS – PLS	28
2.8. ANÁLISE DISCRIMINANTE POR MÍNIMOS QUADRADOS PARCIAIS (PLS-DA) 31	
2.9. MODELAGEM INDEPENDENTE POR ANALOGIA DE CLASSE – SIMCA 31	
3. OBJETIVOS	34
3.1. OBJETIVO GERAL.....	34
3.2. OBJETIVOS ESPECÍFICOS	34
4. ENSAIOS MICROBIOLÓGICOS	36
4.1. MATERIAL E MÉTODOS	37
4.1.1. Programas computacionais.....	37
4.1.2. Reagentes e insumos	37
4.1.3. Teste de Halo de Inibição.....	38
4.1.4. Challenge Test	39
4.2. RESULTADOS E DISCUSSÃO	40
4.2.1. Resultados dos testes de <i>Challenge Test</i> e Halo de Inibição	40
4.2.2. Pré-tratamento dos dados para análise de componentes principais	

4.2.3.	Análise de Componentes Principais.....	41
4.2.4.	Análise dos <i>loadings</i> da PCA.....	43
4.2.1.	Análise de árvore de decisão	44
4.3.	CONSIDERAÇÕES FINAIS E CONCLUSÕES	45
5.	CAPRILATO DE POLIGLICERILA	47
5.1.	MATERIAL E MÉTODOS	48
5.1.1.	Programas computacionais.....	48
5.1.2.	Reagentes e insumos	48
5.1.3.	Desenvolvimentos de modelos de calibração multivariada	50
5.2.	RESULTADOS E DISCUSSÃO	51
5.2.1.	Quantificação do Caprilato de Poliglicerila através da regressão por mínimos quadrados parciais (PLS)	51
5.3.	CONSIDERAÇÕES FINAIS E CONCLUSÕES	56
6.	SISTEMA RECOMENDADOR DE EMOLIENTES	58
6.1.	MATERIAL E MÉTODOS	59
6.1.1.	Programas computacionais.....	59
6.1.2.	Reagentes e insumos	59
6.1.3.	Análises sensoriais e físico-químicas.....	61
6.2.	RESULTADOS E DISCUSSÃO	62
6.2.1.	Pré-tratamento dos dados para análise de componentes principais 62	
6.2.1.	Análise Hierárquica de Agrupamentos (HCA)	64
6.2.2.	Análise por componentes principais.....	65
6.3.	CONSIDERAÇÕES FINAIS E CONCLUSÕES	68
7.	SABONETES EM BARRA	70
7.1.	MATERIAL E MÉTODOS	72
7.1.1.	Equipamentos	72
7.1.2.	Programas computacionais.....	72

7.1.3.	Reagentes e insumos	72
7.1.4.	Classificação das amostras e desenvolvimentos de modelos de calibração multivariada.....	73
7.2.	RESULTADOS E DISCUSSÃO	74
7.2.1.	Pré-tratamento dos dados para análise de componentes principais	74
7.2.2.	Análise por componentes principais.....	75
7.2.3.	Análise discriminante via mínimos quadrados parciais (PLS-DA)	78
7.2.4.	SIMCA - Modelagem Independente por Analogia de Classe	83
7.3.	CONSIDERAÇÕES FINAIS E CONCLUSÕES	88
8.	CONCLUSÕES GERAIS	89
	REFERÊNCIAS	90

1. INTRODUÇÃO

Desde a década de 1990, ferramentas quimiométricas são utilizadas nas diversas áreas da química, principalmente com o objetivo de explorar toda a informação proporcionada pelos métodos instrumentais de análise. Desta forma, técnicas de análise exploratória de dados (ex. Análise por Componentes Principais, PCA) e de desenvolvimento de modelos multivariados (ex. Regressão por Mínimos Quadrados Parciais, PLS) tem contribuído com a resolução de inúmeros problemas analíticos clássicos, assim como, tem servido de base para aplicações menos convencionais, como a classificação arqueológica (CARRERO et al., 2010).

Em função da indústria cosmética se caracterizar pela fabricação de uma grande diversidade de produtos, a necessidade de diversas rotinas de controle de qualidade é bastante evidente. Assim, o controle de qualidade de matérias-primas e produtos acabados normalmente envolve o uso de técnicas analíticas clássicas, como a titulometria, ferramentas instrumentais de análise, incluindo métodos espectroscópicos e cromatográficos, assim como alternativas menos objetivas, como a análise sensorial.

Em razão das numerosas necessidades deste segmento industrial e da usual disponibilidade de ferramentas instrumentais de análise, o uso de técnicas quimiométricas poderia ser de grande valor para o desenvolvimento de rotinas de controle de qualidade, agilizando as respostas e, conseqüentemente, a tomada de decisões. Entretanto, poucos trabalhos relatam este tipo associação, o que faz com que muitas possibilidades de aplicação possam ser estudadas.

Em função destes argumentos, o principal objetivo deste trabalho consiste em avaliar a potencialidade do ferramental quimiométrico no campo de pesquisa, desenvolvimento, controle de qualidade de produtos de higiene pessoal e cosméticos.

2. REVISÃO DA LITERATURA

2.1. COSMÉTICOS: HISTÓRIA E PANORAMA ECONÔMICO

Os cuidados envolvendo higiene e beleza tem acompanhado a humanidade desde as primeiras civilizações. Há 30 mil anos, as pessoas pintavam o corpo usando materiais simples como terra, seiva de árvores e folhas esmagadas. Recentes escavações arqueológicas na área da antiga Mesopotâmia encontraram placas de argila trazendo instruções sobre asseio corporal. Contudo, os egípcios foram os primeiros a fazer uso de cosméticos em grande escala. Há milhares de anos eles já usavam óleo de castor como bálsamo protetor e misturavam argila, cinzas e substâncias perfumadas para se banhar. Como maquiagem, usavam uma série de pigmentos, como o khol (minério de antimônio ou manganês); o verde de malaquita, um minério de cobre, e o cinabre, um minério de sulfeto de mercúrio, para pintar os olhos e a face (TREVISAN; MENDA, 2011).

Na Europa do século XVI, motivadas pelo ideal elisabetano de alvura, as mulheres clareavam a pele usando tinta branca à base de chumbo. Já os cabelos eram clareados com uma mistura de enxofre negro, alume e mel sob exposição ao sol. Segundo orientação dos médicos da época, a maioria da população tomava apenas um ou dois banhos por ano. Foi nesse contexto que a perfumaria ganhou força, principalmente na França, adquirindo grande importância para a economia francesa durante o reinado de Luiz XIV (1638-1715). Contudo, o salto inicial da perfumaria ocorreu em 1725 quando o italiano Giovanni Maria Farina se estabeleceu em Colônia, na Alemanha, onde criou a mais antiga casa de fragrâncias do mundo. Seu principal produto foi o perfume que chamou de *Eau de Cologne* em homenagem à cidade que o acolhera. O perfume de Farina se tornou o preferido da nobreza europeia do século XVIII, foi copiado mundo afora, e a denominação água de colônia virou sinônimo de perfume a partir daí (TREVISAN; MENDA, 2011).

No século XX, com o maior acesso da população em geral à cultura e serviços, assim como, a evolução da indústria cinematográfica e de moda, os cosméticos se tornaram um ponto central na emancipação feminina. Balé, cinema

e desfiles de moda se estabeleceram como ditadores de tendências mundiais e a indústria da beleza observou o surgimento de boa parte dos produtos que hoje definem o setor, por exemplo esmalte para unhas, protetor solar, etc. (L'ORÉAL GROUP, 2014).

Em pleno século XXI, a indústria brasileira de Produtos de Higiene Pessoal, Perfumaria e Cosméticos apresentou, ao longo dos últimos anos, crescimento bem mais vigoroso que o restante da indústria: 10,0 % ao ano de crescimento médio no setor contra 3,0 % ao ano do Produto Interno Bruto nacional e 2,2 % ao ano da Indústria Geral (ABIHPEC, 2014). Estudos recentes demonstraram que o uso de cosméticos está intimamente relacionado ao sucesso pessoal e profissional em diversas áreas, principalmente na área de serviços. (GUÉGUEN; JACOB, 2012).

2.2. CONTROLE DE QUALIDADE NA INDÚSTRIA COSMÉTICA

Um dos maiores desafios da indústria moderna consiste em garantir a satisfação e proteção do consumidor. Empresas de bens e serviços têm investido em tecnologia e pessoal qualificado, visando fornecer soluções que se adequem às exigências cada vez mais rigorosas de consumidores e órgãos regulatórios.

Os setores mais fortemente regulamentados são o médico e o farmacêutico, devido ao seu impacto direto nas questões relacionadas à saúde. Entretanto, as agências reguladoras têm aumentado a pressão sobre setores como o de alimentos e cosméticos.

O Controle de Qualidade é definido como o conjunto de atividades e processos destinados a verificar e assegurar que o produto não seja disponibilizado para uso e venda até que se cumpra com a qualidade preestabelecida. A verificação da conformidade das especificações, mais do que um requisito regulatório, tem como objetivo garantir a qualidade, segurança e eficácia do produto comercializado através de ensaios físicos, químicos e microbiológicos das matérias primas, embalagens e produtos acabados (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2008).

Um processo adequado de Controle da Qualidade requer informações provenientes de fontes diversas e não acaba após a liberação do produto para venda. Os dados podem ser obtidos através de diferentes estudos (toxicologia, sensibilidade, reações alérgicas, etc.). Às vezes, a vigilância dos cosméticos em uso é feita para detectar possíveis efeitos colaterais (FERNÁNDEZ DE CÓRDOVA MANENT; GONZÁLEZ ABELLÁN, 2007).

Os testes de controle de qualidade são aplicados a matérias primas e produtos acabados. Boa parte dos ensaios de controle de matérias primas consiste de análises organolépticas, como aspecto, cor e odor (AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA, 2008). Estes procedimentos dependem do treinamento de profissionais para identificar alterações e comparar amostras com padrões, o que tende a ser caro, demorado e passível de influência da subjetividade sensorial de cada indivíduo.

As ferramentas físico-químicas fundamentadas em análise por via úmida e as análises cromatográficas, em fase líquida e gasosa, constituem o coração analítico do controle de qualidade da indústria cosmética, tanto para análise de matérias primas quanto de produtos acabados. Todavia, o potencial dessas ferramentas ainda não foi totalmente explorado, uma vez que usualmente menos de 1 % da informação produzida é de fato explorada.

Aliados ao ferramental descrito, os ensaios espectroscópicos constituem o estado da arte em termos de análise de matrizes complexas e – aliados as ferramentas estatísticas – apresentam altíssimo potencial para impulsionar o desenvolvimento científico e tecnológico dentro da indústria cosmética, como por exemplo, na otimização de formulações e no desenvolvimento de novos métodos de determinação de compostos marcadores da degradação de produtos, visando a avaliação do tempo de vida em prateleira.

2.3. ESPECTROSCOPIA NO INFRAVERMELHO

As técnicas espectroscópicas são baseadas na interação da radiação eletromagnética com os diferentes analitos, sendo que as técnicas moleculares

mais difundidas são as de UV-Vis, fluorescência e infravermelho. A região do infravermelho engloba radiações com número de onda entre 12.800 cm^{-1} e 10 cm^{-1} . O espectro infravermelho pode ser dividido em infravermelho próximo (NIR, do inglês *Near Infrared*), médio (MID, do inglês, *Middle Infrared*) e distante (FAR, do inglês, *Far Infrared*) conforme evidencia a TABELA 1 (HOLLER, F. J.; SKOOG, D. A.; CROUCH, 2009).

TABELA 1 - REGIÕES ESPECTRAIS NO INFRAVERMELHO

Região Espectral	Número de onda (cm^{-1})	Comprimento de onda (nm)	Frequência (Hz)
Próximo (NIR)	12.800 a 4.000	780 a 2.500	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Médio (MID)	4.000 a 200	2.500 a 50.000	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Distante (FAR)	200 a 10	50.000 a 1.000.000	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

Fonte: (HOLLER, F. J.; SKOOG, D. A.; CROUCH, 2009)

Nas últimas décadas tem crescido o alcance das técnicas baseadas na espectroscopia no infravermelho (KHANMOHAMMADI *et al.*, 2012), principalmente como alternativa a métodos destrutivos ou que necessitem de etapas de extração, representando uma opção mais aderente aos princípios da chamada química verde, ou seja, com ausência do uso de solventes e geração mínima de resíduos (KHANMOHAMMADI; GARMARUDI, 2011).

Considerando o trabalho com amostras de difícil coleta ou de alto valor agregado, a técnica de ATR (Reflectância Total Atenuada, do inglês *Attenuated Total Reflectance*) é uma das técnicas baseadas na reflexão da radiação eletromagnética que tem recebido mais atenção devido à sua versatilidade e cujo uso vem aumentando ano a ano. (DECONINCK *et al.*, 2014).

Uma das características das técnicas espectroscópicas é o grande volume de dados gerados, usualmente com centenas ou milhares de dimensões (comprimentos de onda, valores de frequência, etc.). Nesse sentido, técnicas quimiométricas aparecem como um ferramental útil para se tirar o maior proveito dos dados coletados.

2.4. QUIMIOMETRIA

A quimiometria é a área da química que se dedica ao estudo e aplicação de ferramentas matemáticas e estatísticas para planejar experimentos que possam gerar produtos e processos com as características desejáveis e/ou para obter o máximo de informação a partir de um conjunto de dados (SILVA, 2011). A quimiometria vem sendo utilizada com sucesso, tanto na indústria como na área acadêmica, onde o desenvolvimento de diversas ferramentas para tratamento de dados pode ser utilizado em diversas áreas, como otimização de processos, classificação e construção de modelos de regressão.

Durante as últimas décadas, a calibração multivariada tem sido empregada em diversos campos, como por exemplo na caracterização de derivados do petróleo (KHANMOHAMMADI *et al.*, 2012), alimentos (OLIVERI; DOWNEY, 2012), fármacos (RAJALAHTI; KVALHEIM, 2011), agricultura (KANIU; ANGEYO, 2015) e mineração (SMOLIŃSKI *et al.*, 2016). Entretanto, apesar do enorme potencial analítico, as aplicações dentro da indústria cosmética ainda estão restritas a um número reduzido de publicações, com foco principalmente na determinação de compostos em matrizes complexas, como por exemplo, agentes clareadores ilegais em cosméticos (DECONINCK *et al.*, 2014) e na determinação simultânea de agentes fixadores em perfumes (NIE *et al.*, 2008). Outra aplicação em alta é a detecção de falsificações e adulterações, como – por exemplo – análise da origem do óleo essencial da *Aniba rosaeodora* (ALMEIDA *et al.*, 2013) ou adulteração de pasta de dente (LÓPEZ-SÁNCHEZ *et al.*, 2008). Por fim, nos últimos anos têm surgido aplicações que merecem destaque, tais como, a caracterização forense de manchas de batom (LÓPEZ-LÓPEZ *et al.*, 2014) e a previsão de propriedades sensoriais a partir de dados instrumentais (GILBERT *et al.*, 2013).

2.5. ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)

A HCA (do inglês *hierarchical cluster analysis*) se baseia no cálculo das distâncias entre pontos no sistema multivariado de variáveis que representa todas as amostras. Os resultados da HCA são apresentados através de uma representação gráfica, a qual é chamada dendrograma. No procedimento de construção desse tipo de diagrama, cada amostra é agrupada com a amostra mais próxima e assim sucessivamente, até que os agrupamentos (*clusters*) vão também se unindo em um único conjunto contendo todas as amostras. A semelhança ou diferença entre as propriedades físicas ou químicas das amostras é representada através da proximidade com que as mesmas se encontram no dendrograma (NASCIMENTO *et al.*, 2010).

Uma vez que o funcionamento da HCA se baseia no conceito de similaridade (UARROTA *et al.*, 2014), caso duas amostras estejam próximas uma da outra no espaço multidimensional, estas mesmas amostras também devem estar matematicamente próximas no dendrograma.

Para calcular as distâncias entre as amostras, primeiro é necessário definir o tipo de distância empregado, Mahalanobis, Manhattan, Euclidiana, etc. (OLIVERI; DOWNEY, 2012). Neste trabalho, será empregada a distância Euclidiana, a qual pode ser vista de maneira trivial como a distância entre dois pontos num plano, como demonstrado pela Equação 5.

$$d_{ab} = \left(\sum_{j=1}^m (x_{aj} - x_{bj})^2 \right)^{1/2} \quad (5)$$

Em seguida estes valores de distância são convertidos para uma mesma escala de similaridade, a qual é calculada segundo a Equação 6, na qual, S_{ab} e d_{ab} são, respectivamente, a similaridade e a distância entre as amostras a e b. Da mesma maneira, d_{max} é a maior distância entre todas as amostras (MORGANO *et al.*, 1999).

$$S_{ab} = 1 - \left\{ \frac{d_{ab}}{d_{max}} \right\} \quad (6)$$

Conforme demonstrado no dendrograma exemplificativo apresentado na FIGURA 1, trabalha-se com uma escala entre zero (amostras sem similaridade) e um (amostras idênticas). Após determinados os valores de similaridade, as duas amostras mais próximas entre si são conectadas de maneira a formar um agrupamento. Este processo é repetido tantas vezes quanto for necessário até que todas as amostras estejam conectadas formando um único agrupamento. Considerando que ao longo do processo as amostras e agrupamentos podem ser conectados entre si de acordo com a proximidade, deve-se definir os conceitos de distância amostra/grupo e grupo/grupo. Para tanto, vários métodos podem ser utilizados, sendo que os mais comumente empregados são: conexão pelo vizinho mais próximo, conexão pelo vizinho mais distante e conexão pela média.

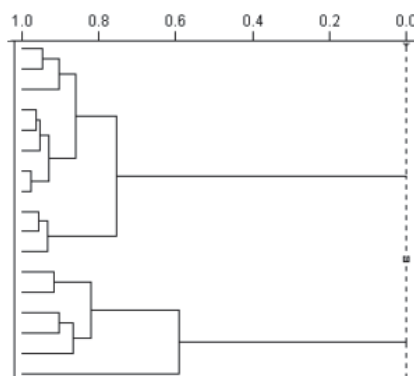


FIGURA 1 – EXEMPLO DE REPRESENTAÇÃO DE AGRUPAMENTOS VIA DENDROGRAMA
FONTE: (MORGANO et al., 1999)

ANÁLISE POR COMPONENTES PRINCIPAIS – PCA

A Análise por Componentes Principais (PCA, do inglês: *Principal Component Analysis*) é uma técnica de compressão que permite reduzir a dimensionalidade de uma matriz de dados (**X**), utilizando um pequeno número de novas variáveis, denominadas componentes principais (PCs), que surgem da combinação linear das variáveis originais.

O objetivo principal deste processo de compressão consiste em expressar a maior parte da variância contida na matriz de variáveis originais (\mathbf{X}), utilizando um pequeno número de PCs, as quais, pelo princípio da ortogonalidade, representam informações complementares.

A ilustração apresentada na FIGURA 2a representa um conjunto de dados que se distribui em um espaço tridimensional. Na FIGURA 2b representa-se a criação de três componentes principais, as quais, por definição, são enumeradas de acordo com a quantidade de variância capturada. Assim, a PC1 explica grande parte da variância dos dados originais, sendo complementada pela PC2, que, pelo conceito de ortogonalidade, representa a variância que não pode ser explicada pela primeira PC. De acordo com a ilustração, a terceira PC representa muito pouco da variância total, podendo, eventualmente, ser desconsiderada na interpretação dos dados. Desta forma, grande parte da variância pode ser adequadamente representada por apenas duas PC, o que resulta em uma redução de dimensionalidade e facilita a visualização dos dados, tal como mostrado na FIGURA 2c.

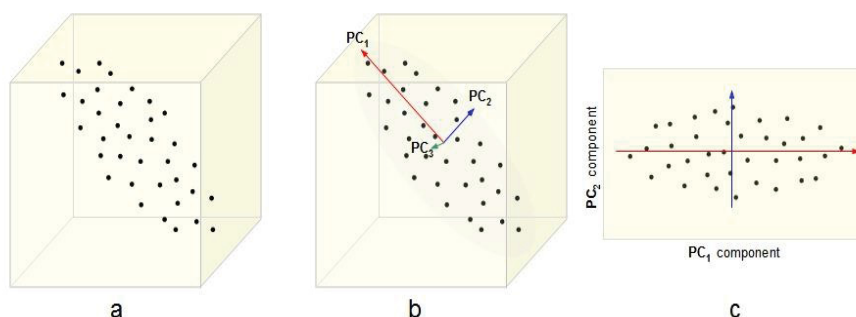


FIGURA 2 - ILUSTRAÇÃO DO PROCESSO DE REDUÇÃO DE DIMENSIONALIDADE. a) CONJUNTO TRIDIMENSIONAL DE DADOS; b) PROPOSTA DE 3 COMPONENTES PRINCIPAIS; E c) PROJEÇÃO DOS ELEMENTOS ORIGINAIS EM FUNÇÃO DE DUAS PCS (KAVRAKI, 2007)

Na PCA, cada nova componente principal é decomposta em duas novas matrizes, uma de *scores* e outra de *loadings* (pesos), o que permite representar a matriz original \mathbf{X} de acordo com a expressão apresentada na Equação 1. Os *scores* representam as coordenadas das amostras nos novos eixos (PCs), enquanto que os *loadings* representam a informação do quanto cada variável original contribui para a formação de cada novo eixo (PCs). Desta forma, os *scores* expressam as

relações entre as amostras, enquanto os *loadings* mostram as relações entre as variáveis originais, como pode ser observado na FIGURA 3.

$$X = \sum_{i=1}^A T * P^t \quad (1)$$

Onde: T representa os *scores* e P os *loadings*.

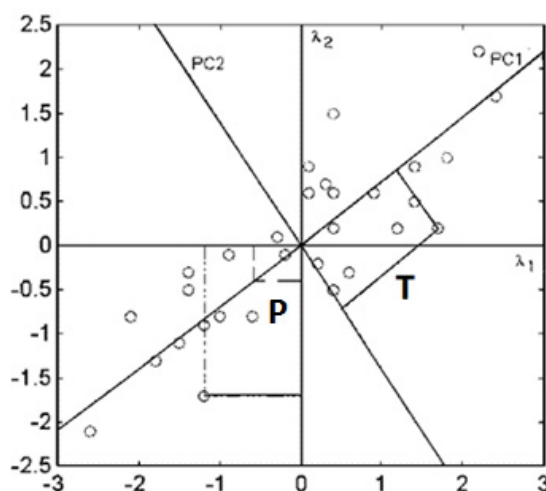


FIGURA 3 - CONJUNTO DE DADOS BIDIMENSIONAIS MOSTRANDO OS EIXOS DAS VARIÁVEIS ORIGINAIS (λ_1 e λ_2), DAS COMPONENTES PRINCIPAIS (PC1 E PC2) E A REPRESENTAÇÃO DE SCORES (T) E LOADINGS (P) (FERREIRA *et al.*, 1999)

A principal vantagem obtida pela aplicação da PCA é a possibilidade de visualizar de forma gráfica a informação presente em uma matriz de dados de grandes dimensões. O algoritmo, ao reduzir a dimensionalidade da matriz de dados, seleciona com base nas instruções do usuário as parcelas da informação que apresentam variância significativa no sistema, possibilitando observar as similaridades e diferenças entre as amostras. Adicionalmente, é possível inferir quais parcelas dos dados originais explicam os agrupamentos das amostras de acordo com semelhanças das mesmas. A análise conjunta dos resultados (*scores* e *loadings*) possibilita a visualização de padrões e correlações cuja existência seria mais difícil de ser vislumbrada sem o uso de ferramentas multivariadas.

As menções na literatura envolvendo o uso de PCA para solucionar problemas dentro do ambiente cosmético, embora ainda pouco numerosas, apontam para o potencial da técnica na evolução de plataformas tecnológicas

dentro do ambiente cosmético. Casos como a determinação de ativos (HANGANU *et al.*, 2012) e avaliação sensorial (WORTEL; WIECHERS, 2000) ilustram essa tendência e exemplificam os caminhos ainda pouco explorados da análise quimiométrica para a indústria de beleza.

2.6. REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS – PLS

As ferramentas de regressão têm como objetivo estabelecer, sobre condições específicas, uma relação matemática entre os dados instrumentais e os valores relacionados às propriedades físicas ou químicas de interesse.

O universo desses métodos se divide entre os univariados e os multivariados. Os métodos univariados consistem na correlação de uma única resposta instrumental para cada amostra e a propriedade de interesse. Enquanto os métodos multivariados permitem a utilização de múltiplas respostas para serem relacionadas às propriedades das amostras, fato que permite a obtenção da correlação desejada.

Tradicionalmente em análises quantitativas, por exemplo: cromatográficas ou espectroscópicas, determina-se o teor de certos analitos através da construção de curvas de calibração convencionais. Este sistema baseia-se na criação de equações lineares que relacionam a resposta instrumental de uma determinada amostra à concentração da substância de interesse. Este tipo de análise ainda é o cerne dos programas de controle da qualidade na maioria das empresas. Uma característica desse tipo de sistema é que ele depende apenas de uma única resposta (área do pico em determinado tempo de retenção, ou absorbância em um determinado comprimento de onda), sendo que a maioria das análises modernas, como por exemplo, as análises espectroscópicas (NIR, UV-Vis, etc.), fornece um número muito maior de informações que geralmente não é aproveitado.

Nesse contexto, os métodos de calibração multivariada, como a Regressão por Mínimos Quadrados Parciais (PLS), aparecem como uma oportunidade de obtenção de resultados mais robustos e abrangentes, sendo possível empregar um número elevado de valores de resposta instrumental para cada uma das amostras. A técnica de PLS foi desenvolvida entre as décadas de 1970 e 1980 por Herman

Wold (WOLD, 1982), envolvendo trabalhos em econometria. Atualmente, é um dos procedimentos de regressão mais utilizados em calibração multivariada (KHANI *et al.*, 2014).

Este método de regressão pode ser dividido em PLS1 (modelo de regressão otimizado e obtido para uma variável dependente de cada vez) e PLS2 (modelo de regressão otimizado simultaneamente para duas ou mais variáveis dependentes), sendo que, em ambos os casos, trata-se de uma análise robusta, ou seja, os resultados não se alteram consideravelmente com a inclusão de novas amostras no conjunto de calibração (FERREIRA *et al.*, 1999).

Em termos gerais, pode-se definir que a calibração multivariada visa relacionar a matriz de dados experimentais (**X**), denominada de matriz de variáveis independentes (espectros, dados físico-químicos, etc.), com as propriedades de interesse (matriz **Y**), formada pelas variáveis dependentes: concentração, eficácia, etc. (FERREIRA *et al.*, 1999)

Inicialmente, cada uma destas matrizes é submetida a um processo de redução de dimensionalidade, de maneira análoga à PCA, sendo que cada matriz pode ser representada por uma matriz de *scores* (**T** e **U**) e *loadings* (**P** e **Q**), tal como pela Equação 2 e pela Equação 3.

$$X = T * P^t + E \quad (2)$$

$$Y = U * Q^t + R \quad (3)$$

Finalmente, o modelo é desenvolvido correlacionando-se os *scores* da matriz **X** (**T**) com os da matriz **Y** (**U**), obtendo-se um vetor de regressão (**b**) que permite a previsão de **Y** a partir de **X** (Equação 4).

$$U = b * T \quad (4)$$

Visando-se obter uma melhor correlação entre as variáveis, por meio da captação da maior quantidade possível de informação presente nos dados (variância), faz-se necessária a manipulação das matrizes **X** e **Y** de maneira simultânea. No PLS, esse tipo de resultado é obtido através da rotação das componentes principais, levando a uma pequena distorção da ortogonalidade.

Devido a isso, as componentes principais passam a ser chamadas de variáveis latentes (SILVA, 2011).

Durante o procedimento de obtenção do vetor de coeficientes de regressão que melhor represente a correlação desejada, dois fatores devem ser considerados: o número de variáveis latentes a serem mantidos no modelo e a ausência de amostras anômalas no conjunto de calibração. Para auxiliar nestas decisões, procedimentos de validação cruzada podem ser utilizados, de maneira que a calibração é repetida n vezes (n = número de amostras do conjunto de calibração), sendo que em cada modelagem uma ou mais amostras do conjunto de calibração são retiradas e utilizadas como elementos de previsão, empregando-se um número crescente de variáveis latentes. Assim, calcula-se a raiz quadrada do erro quadrático médio de validação cruzada (RMSECV) para determinar o número de variáveis latentes que gera o menor erro de previsão. O modelo otimizado pode ser então empregado, para testar a sua capacidade preditiva. O processo de validação externa é realizado através da aplicação do modelo em amostras (não usadas na etapa de calibração) visando determinar a concentração da(s) espécie(s) de interesse. Em seguida, os resultados fornecidos pelo modelo são comparados com os valores de referência para estabelecer se ele de fato reflete o comportamento do(s) analito(s). Embora as rotinas de análise multivariada tenham mostrado um ótimo desempenho em muitas aplicações de relevância (RAJALAHTI; KVALHEIM, 2011); (GENDRIN *et al.*, 2008); (ESCANDAR *et al.*, 2006), a sua utilização na área da produção de cosméticos é praticamente inexistente. Assim, dos mais de mil trabalhos publicados entre 1990 e 2014 na área (**Fonte:** Sciencedirect, **Palavras Chave:** *Cosmetic and Analysis*), apenas 12 envolvem algum tipo de análise multivariada. Dentro deste contexto, destaca-se o recente trabalho de GILBERT *et al.* (2013), que propõe o uso de PLS para desenvolvimento de modelos multivariados que permitam prever resultados da análise sensorial de textura, a partir da caracterização instrumental de emulsões cosméticas.

Outro trabalho que merece destaque foi o de KULIGOWSKI *et al.* (2012), que desenvolveu um método rápido e barato para controle da produção de depiladores a base de açúcar, utilizando espectroscopia no infravermelho (ATR-FTIR) e regressão multivariada (PLS1). Foram construídos modelos para determinação de glicose, frutose e maltose em depiladores, sem qualquer tipo de

preparação de amostra. A análise revelou uma mudança na concentração dos açúcares durante o processo de formulação do produto.

2.7. ANÁLISE DISCRIMINANTE POR MÍNIMOS QUADRADOS PARCIAIS (PLS-DA)

No ambiente Matlab, o funcionamento do PLS-DA se dá baseado no arcabouço matemático descrito anteriormente para a técnica de PLS, com a particularidade de que nesse caso a matriz Y é composta por valores numéricos inteiros referentes cada um à respectiva classe (em um exemplo com duas classes, a matriz Y poderia ser composta por valores de 1 para as amostras pertencentes à primeira classe e 2 para amostras pertencentes à segunda classe). Tanto a matriz X quanto a matriz Y são então decompostas no produto de duas matrizes (*scores* e *loadings*) e um vetor de coeficientes de regressão que melhor correlacione os *scores* de ambas as matrizes é obtido. Para a etapa de classificação, compara-se o valor previsto pelo modelo com os valores designados para cada classe e, quanto mais próximo um elemento de uma determinada coluna em Y está do valor da classe, com certo intervalo de confiança (usualmente 95 %), maior a probabilidade de referida amostra pertencer à dita classe (BERRUETA *et al.*, 2007).

2.8. MODELAGEM INDEPENDENTE POR ANALOGIA DE CLASSE – SIMCA

O método de classificação SIMCA tem seu funcionamento baseado no reconhecimento de padrões e similaridade. Ele analisa a classificação fornecida para um dado conjunto de treinamento, estima um grau de confiança da classificação e prevê a qual grupo pertence cada nova amostra introduzida. Um resultado possível é que as amostras testadas sejam classificadas como não pertencentes a nenhum dos grupos presentes no conjunto de calibração (DAGO MORALES *et al.*, 2008).

Quando um modelo SIMCA classifica uma amostra como pertencente a uma classe X, significa que os espectros ou conjuntos de variáveis desta amostra são semelhantes aos espectros ou variáveis das amostras utilizadas para a construção

da classe X do modelo. A construção de modelos SIMCA se dá primeiramente pela determinação, de maneira independente para cada classe, do número de componentes principais (PCs) necessários para descrever cada conjunto de amostras (classes). Em seguida, procede-se com a construção de hipercaixas para cada uma das classes de acordo com o número otimizado de PCs para cada uma delas, onde as fronteiras das mesmas são definidas em função do desvio padrão dos *scores* em cada componente principal com um dado nível de confiança (geralmente 95 % de confiança). Nesta etapa é possível identificar a presença de amostras anômalas no conjunto de treinamento, determinando-se a variância residual para cada amostra na classe (Q , representa a distância ortogonal entre a amostra e o plano definido pelas componentes principais) e a variância residual total (S_o , parâmetro que representa o desvio padrão dentro da classe, ou seja, a compactação das amostras na classe). Em seguida, é aplicado um teste F (Q^2/S_o^2) para verificar a presença de anomalias de acordo com o F_{crit} para um determinado número de graus de liberdade dentro de certo nível de probabilidade. Se $Q^2 \gg S_o^2$, indica que a amostra está distante do modelo e pode ser considerada uma amostra anômala nesta classe, mas ela pode pertencer a uma outra classe que está sendo modelada.

Um dos principais critérios para definição do poder discriminatório de um modelo SIMCA é a distância entre as hipercaixas, quanto maior esta distância, menor será a chance de que haja sobreposições entre as mesmas, o que poderia acarretar um possível erro de classificação do modelo. A classificação de uma nova amostra é realizada determinando-se a projeção da mesma no espaço dos *scores* e a sua distância das fronteiras de cada uma das classes em questão. Caso o *score* calculado para a nova amostra caia fora dos limites das hipercaixas, ela será classificada como não pertencente a nenhuma das classes, mas é possível calcular sua distância até a fronteira de cada uma das classes para identificar com qual delas ela mais se assemelha.

Além dos resultados da classificação para cada amostra, o SIMCA disponibilizado no pacote PLS Toolbox operando no ambiente Matlab fornece para cada classe os gráficos de PRESS (soma quadrática dos erros de previsão), autovalores da variância acumulada dos conjuntos X e Y; e um gráfico de Q vs T^2 , parâmetros estes que representam respectivamente os resíduos e a distância de cada amostra em relação ao centro da hipercaixa. Conforme demonstrado na

FIGURA 4, o parâmetro T^2 de Hotelling pode ser visto como a distância da projeção da amostra no subespaço k -dimensional até o centróide do subespaço (intersecção das PCs da respectiva classe). O parâmetro de resíduos Q é a distância entre a amostra no n -espaço e a sua projeção no subespaço k -dimensional do modelo.

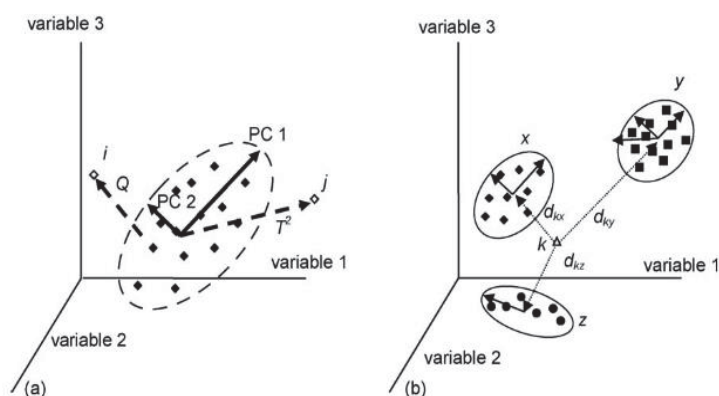


FIGURA 4 – (a) REPRESENTAÇÃO GRÁFICA DOS PARÂMETROS T^2 DE HOTELLING E Q RESIDUAL PARA TRÊS VARIÁVEIS E DUAS PCs. (b) MODELO SIMCA PARA TRÊS CLASSES, SENDO QUE CADA CLASSE APRESENTA UM NÚMERO DIFERENTE DE COMPONENTES PRINCIPAIS E AMOSTRAS DE CALIBRAÇÃO. AS DISTÂNCIAS DO PONTO k PARA CADA UMA DAS CLASSES TAMBÉM ESTÁ REPRESENTADA (HALL; KENNY, 2007)

A amostra é definida como pertencente à classe modelada se tanto o valor de Q quanto o de T^2 for menor que os valores críticos. Os resultados são estabelecidos para cada classe separadamente. Caso uma amostra testada apresente valor de distância superior aos valores estipulados para cada classe, a amostra é classificada como não pertencente a nenhuma das classes. O padrão para este tipo de análise classificatória é o emprego de 95 % de confiabilidade.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Estudar algumas aplicações do ferramental quimiométrico no campo de pesquisa, desenvolvimento e controle de qualidade de produtos de higiene pessoal, perfumaria e cosméticos.

3.2. OBJETIVOS ESPECÍFICOS

Dentre os objetivos específicos destacam-se:

- Desenvolver um método baseado na técnica de espectroscopia na região do infravermelho por ATR que identifique a massa base empregada em sabonetes comerciais (presença de massa base animal, vegetal ou sintética), visando pesquisa e desenvolvimento e comprovação de apelos veganos nas embalagens.
- Verificar a capacidade do Teste de Halo de Inibição de microrganismos na previsão dos resultados do *Challenge Test* (estudo da efetividade de sistemas conservantes em fórmulas cosméticas), a fim de reduzir o tempo de desenvolvimento e otimizar o trabalho de *screening* de novos compostos.
- Desenvolver um método para a determinação do conservante Caprilato de Poliglicerila em formulações cosméticas fazendo uso da técnica de ATR associada a ferramentas multivariadas (PLS).
- Estabelecer um sistema multivariado de agrupamentos de emolientes de acordo com suas características físico-químicas e relacioná-los às respectivas aplicações cosméticas empregando as metodologias de HCA e PCA.

Capítulo I – ENSAIOS MICROBIOLÓGICOS

4. ENSAIOS MICROBIOLÓGICOS

O chamado *Challenge Test*, é amplamente utilizado na indústria de cosméticos para avaliar a eficácia de sistemas conservantes para uma dada formulação. Este teste consiste na contaminação intencional do produto com os microrganismos e avaliação da carga microbiana em intervalos de tempo pré-definidos (BRASIL, 2012). No entanto, trata-se de um ensaio demorado, levando 28 dias para ser finalizado, o que pode comprometer a triagem de novos sistemas de conservação.

Outro teste microbiológico bastante utilizado é o teste de halo de inibição, o qual envolve a avaliação de um microrganismo num meio de crescimento contra um agente antimicrobiano por um período de tempo determinado. Assim, é possível relacionar o tamanho da zona de inibição do crescimento do microrganismo com a eficácia antimicrobiana, bacteriostática ou fungistática da substância (Pinto *et al.*, 2003) (PINTO *et al.*, 2003). Este é um teste rápido (48 horas de duração) e de relativa simplicidade em termos de implementação.

4.1. MATERIAL E MÉTODOS

4.1.1. Programas computacionais

Para a montagem das matrizes de dados foi utilizado o software Origin Pro 9.1[®] (OriginLab), enquanto que para a elaboração dos modelos foi empregado o pacote PLS-Toolbox 3.0 (Eigenvector Research, Inc.) que opera em ambiente Matlab[®] v.7.0.1 (Math Work Inc.).

Para a implementação do algoritmo de árvore de decisão J48, foi empregado o software Weka 3.0.

4.1.2. Reagentes e insumos

Os padrões e amostras necessários foram proporcionados pelo Grupo Boticário ou adquiridos no mercado em grau de pureza compatível com os objetivos do estudo.

Todos os materiais e vidrarias empregados no trabalho foram previamente calibrados e descontaminados. A TABELA 2 apresenta a lista dos sistemas conservantes estudados.

TABELA 2 – SISTEMAS CONSERVANTES ESTUDADOS

Código	Sistema Conservante (INCI Name)	Origem
A	Ácido levulínico, Levulinato de sódio, Glicerina, Água	Natural
B	Caprilato de glicerila	Natural
C	Undecilenato de glicerila	Natural
D	Extrato da folha de <i>Olea europaea</i>	Natural
E	Sesquicaprilato de Xilitol, Anidroxilitol	Natural
F	Fenilpropanol, Extrato do Fruto de <i>Citrus reticulata</i> , Extrato do fruto de <i>Citrus aurantium amara</i> , Extrato da pele de <i>Citrus sinensis</i>	Natural e Sintético
G	Ácido caprilhidroxâmico, caprilato de glicerila, metilpropanediol	Natural e Sintético
H	Ácido p-anísico	Sintético
I	Benzoato de sódio, Sorbato de potássio	Sintético
J	Carvacrol, p-cimeno, 1,4-cineol, propilenoglicol (Plus)	Sintético
K	Gluconolactona, Benzoato de sódio, Gluconato de cálcio	Sintético
L	Caprilil glicol	Sintético

Os testes microbiológicos foram realizados no laboratório de microbiologia do Grupo Boticário de acordo com metodologias e procedimentos internos.

4.1.3. Teste de Halo de Inibição

Os microrganismos testados neste estudo foram: *Escherichia coli*, *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Candida Albicans* e *Aspergillus niger*.

Após cultivo dos microrganismos em meio de cultura padrão (BARICELLI *et al.*, 2015), discos contendo alíquotas de cada sistema de conservantes foram aplicados ao centro das placas de Petri contendo os fungos ou bactérias mencionados acima. Como resultado, após 48 horas, foram realizadas medidas do halo de inibição em milímetros da atividade microbiológica de cada sistema conservante sobre cada organismo.

Capítulo I – Ensaios microbiológicos

4.1.4. Challenge Test

Para a realização do *Challenge Test*, foi seguida a metodologia descrita por CTFA (2001).

Primeiramente foi preparada uma fórmula base representativa de um creme cosmético padrão para cuidados da pele com a seguinte composição: estearato de glicerila, palmitato de cetila, álcool estearílico, álcool cetílico, etilhexil estearato, cetearato-20, cetearato-5, adipato de dibutila, triglicerídeo caprílico/cáprico, hidroxitolueno butilado (BHT), ácido etilenodiamino tetra-acético (EDTA), xilitol, glicerina, goma xantana, carbomero, água, trietanolamina, propileno glicol, pantenol, ciclometicone, extrato de *Punica granatum*, extrato de semente de *Pimpinella anisum*.

Em seguida, foram adicionados os sistemas conservantes nas concentrações descritas na TABELA 2 à fórmula base. Cada uma dessas amostras foi identificada e armazenada em condições padrão para o teste ($26,0 \pm 2,0$ °C).

Em intervalos de tempo de 2, 7, 14 e 28 dias foram realizadas coletas contendo alíquotas de cada amostra. Cada alíquota foi submetida a uma contaminação intencional seguida de um período de incubação e verificação de eficácia da ação microbicida (pela contagem do número de unidades formadoras de colônia) com cada um dos seguintes microrganismos: *Escherichia coli*, *Staphylococcus aureus*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Burkholderia cepacia*, *Enterobacter cloacae*, *Candida Albicans* e *Aspergillus niger* (CTFA, 2001).

4.2. RESULTADOS E DISCUSSÃO

4.2.1. Resultados dos testes de *Challenge Test* e Halo de Inibição

Na TABELA 3 são apresentados os resultados dos testes microbiológicos realizados. Para cada um dos sistemas conservantes, procurou-se padrões de correlação entre os resultados dos testes de Halo de Inibição e o *Challenge Test*, contudo nenhum padrão óbvio pôde ser discernido.

TABELA 3 – RESULTADOS DOS TESTES MICROBIOLÓGICOS PARA OS SISTEMAS CONSERVANTES

Sistema Conservante	Resultado final <i>Challenge Test</i>	Resultados do Teste de Halo de Inibição (mm)				
		<i>Staphylococcus aureus</i>	<i>Pseudomonas aeruginosa</i>	<i>Escherichia coli</i>	<i>Candida albicans</i>	<i>Aspergillus niger</i>
A	Ineficiente	4	7	7	3	1
B	Ineficiente	5	2	4	10	11
C	Ineficiente	1	0,5	1	6	7
D	Ineficiente	0	0	0	0	0
E	Ineficiente	3	0,5	1	1	4
F	Eficiente	4	3	1	7	15
G	Eficiente	6	1	8	10	15
H	Ineficiente	0,5	0	1	2	0,5
I	Eficiente	10	0,5	5	12	18
J	Ineficiente	18	1	4	15	20
K	Eficiente	8	0,5	2,5	10	16
L	Eficiente	7	3	0	7	15

4.2.2. Pré-tratamento dos dados para análise de componentes principais

A fim de utilizar todo o potencial do reconhecimento de padrões e diferenciação de grupos da técnica de análise de componentes principais (PCA), faz-se necessário a aplicação de um ferramental matemático como filtros e operadores (pré-processamentos) visando salientar as pequenas diferenças nas variáveis independentes (matriz de dados X) e assim, tornar a interpretação dos resultados da PCA mais fácil (KAVRAKI, 2007).

Capítulo I – Ensaio microbiológicos

A amplitude das diferentes variáveis (valores de Halo de Inibição para os diferentes microrganismos) não apresenta a mesma escala, indo de zero a oito para a *Escherichia coli* e de zero a vinte para o *Aspergillus niger*. Essa diferença de escala pode não parecer significativa a primeira vista, mas adiciona variações que não são decorrentes necessariamente da eficácia de cada sistema conservante e sim dos mecanismos de ação intrínsecos de cada microrganismo. Tal comportamento pode levar a uma maior ponderação das variáveis com maior valor numérico (no caso, os valores de halo para o *Aspergillus niger*). Uma forma de contornar esse empecilho, sem que haja perdas de informação significativa dos dados, é realizar um autoescalamento sobre o conjunto original dos dados, de modo que cada variável apresente média zero e variância igual a um. (MOITA NETO; MOITA, 1998).

Nesse contexto, o autoescalamento foi o procedimento selecionado como pré-tratamento para a matriz X de dados.

4.2.3. Análise de Componentes Principais

Depois de realizada a etapa de pré-tratamento dos dados e a fim de selecionar quais resultados do teste de Halo de Inibição melhor serviriam de modelo preditivo para os resultados do *Challenge Test*, foi feito uso da metodologia de identificação de padrões por meio da redução de dimensionalidade através de uma análise de componentes principais (PCA).

Como demonstrado na FIGURA 5, com apenas duas componentes principais (PCs) é possível explicar 84,65 % da variância do sistema. Em outras palavras, a maior parte da informação proveniente de 5 variáveis distintas pode agora ser visualizada em apenas duas dimensões, as quais podem ser representadas em um gráfico 2D de coordenadas cartesianas.

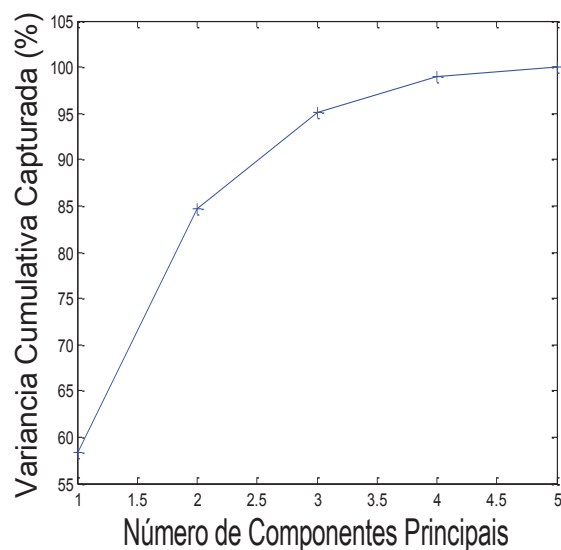


FIGURA 5 – VARIÂNCIA ACUMULADA COM O AUMENTO DO NÚMERO DE COMPONENTES PRINCIPAIS

A FIGURA 6 contém a apresentação gráfica dos *scores* das componentes principais (PCs) 1 e 2 geradas pela PCA conforme descrito. Podem ser observados dois grupos divididos espacialmente, sendo que as amostras com resultado ineficiente no *Challenge Test* apresentam-se agrupadas em sua maioria do lado negativo da PC1 e as com resultado eficiente estão todas agrupadas do lado positivo da PC1.

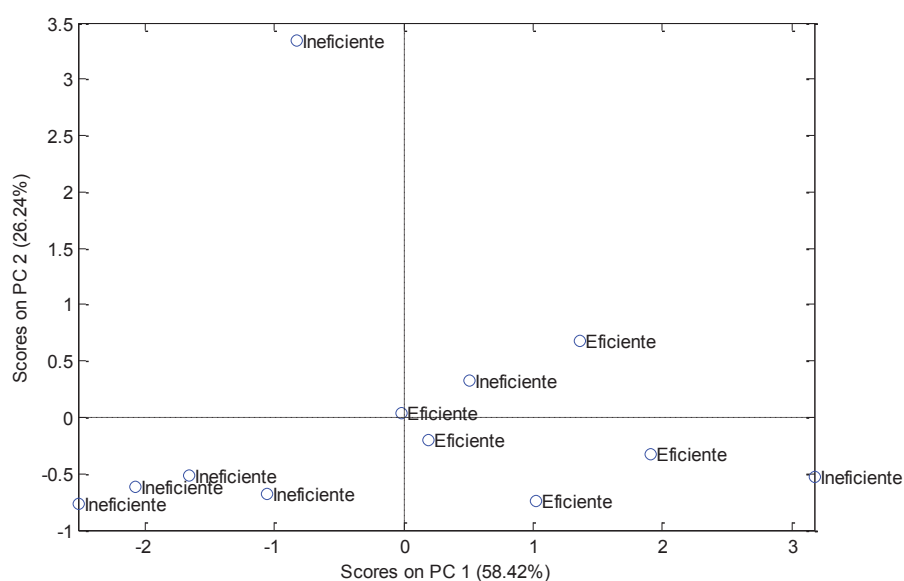


FIGURA 6 – APRESENTAÇÃO GRÁFICAS DAS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS. AMOSTRAS EFICIENTES E INEFICIENTES NO *CHALLENGE TEST*

4.2.4. Análise dos *loadings* da PCA

Visando entender quais microrganismos testados via halo de inibição seriam os melhores candidatos a marcadores preditivos dos resultados do *Challenge Test*, uma análise dos gráficos dos *loadings* (FIGURA 7) das respectivas PCs foi realizada. Uma vez que os *loadings* representam o peso que cada variável original (halo de inibição do conservante sobre cada microrganismo) tem na construção das PCs, é possível observar que os microrganismos *Staphylococcus aureus*, *Candida Albicans* e *Aspergillus niger* foram aqueles que apresentaram maior relevância para a construção da PC1, a qual representa 58,42 % da variância do sistema.

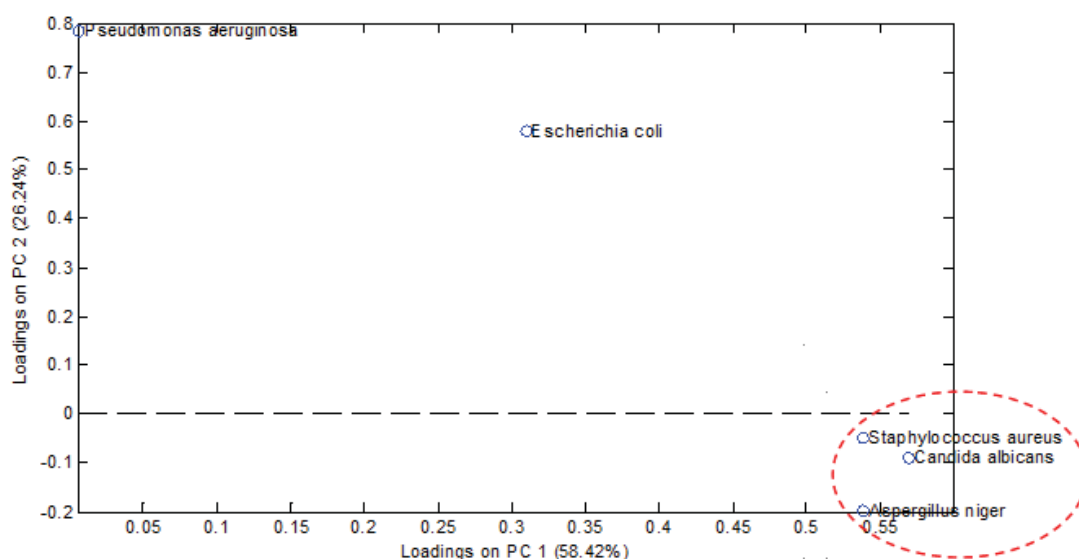


FIGURA 7 - LOADINGS DAS PCS 1 E 2

Depois de realizada a análise exploratória através da aplicação da PCA, o estudo ainda careceu de dados mais objetivos para definição de um método preditivo. Devido a isto, a ferramenta de árvore de decisão foi empregada para a obtenção de informações mais assertivas sobre como prever os resultados do *Challenge Test* (quase um mês de duração) a partir dos resultados do Teste de Halo de Inibição (dois dias).

Capítulo I – Ensaios microbiológicos

4.2.1. Análise de árvore de decisão

Com base nas informações obtidas pela PCA e fazendo uso do algoritmo J48, uma implementação do algoritmo de árvore de decisão C4.5, o qual trabalha construindo uma árvore de decisão (forma gráfica de visualização de dados numéricos), foram analisados os dados desses dois ensaios microbiológicos (HALL *et al.*, 2009). Os resultados são mostrados na FIGURA 8.

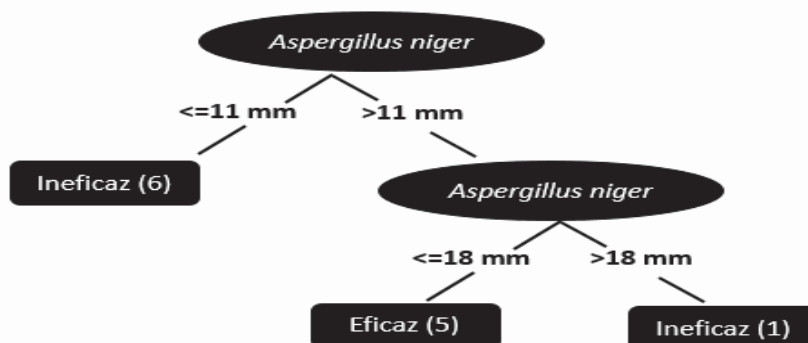


FIGURA 8 – RESULTADOS DA ÁRVORE DE DECISÃO. NÚMERO DE SISTEMAS CONSERVANTES CLASSIFICADOS EM CADA GRUPO CONSTA ENTRE PARENTESSES

O resultado combinado das duas análises indica que o Teste de Halo de Inibição para o *Aspergillus niger* constitui um marcador preditivo adequado para prever os resultados de um sistema conservante frente ao *Challenge Test*. Em termos práticos, para ser considerado eficaz frente ao *Challenge Test*, o halo de inibição para o fungo *Aspergillus niger* deve ser superior a 11 mm. No entanto, se o halo de inibição for maior do que 18 mm, um resultado ineficaz pode ser esperado.

A explicação para o resultado de ineficácia frente ao *Challenge Test* para grandes valores de halo de inibição foi atribuída à maior oleosidade do sistema conservante J (Carvacrol, p-cimeno, 1,4-cineol, propilenoglicol (Plus)), o qual – devido ao baixo coeficiente de solubilização na fórmula cosmética empregada – teria se espalhado por sobre a superfície, gerando altos valores de halo devido à proteção superficial, não apresentando a mesma eficácia quando dissolvido na amostra, como ocorre no *Challenge Test*.

4.3. CONSIDERAÇÕES FINAIS E CONCLUSÕES

Foi desenvolvida uma metodologia baseada no Teste de Halo de Inibição para previsão dos resultados do *Challenge Test* na avaliação de eficácia de sistemas conservantes.

A ferramenta de triagem proposta aponta que o Teste de Halo de Inibição para o fungo *Aspergillus niger* constitui um marcador preditivo adequado para prever os resultados de um sistema conservante frente ao *Challenge Test*, no qual valores entre 11 e 18 mm apontam que um resultado eficaz pode ser esperado.

Apesar do controle microbiológico ser uma ciência complexa e requerer estudos aprofundados visando garantir a segurança dos produtos desenvolvidos, o método de trabalho proposto reduz de 28 para 2 dias o período de comparação de diferentes conservantes para uso em produtos, gerando ganhos econômicos e de tempo na produção de novas fórmulas.

Capítulo II – CAPRILATO DE POLIGLICERILA

5. CAPRILATO DE POLIGLICERILA

Comercializado com o nome de TEGO® Cosmo P 813, o Caprilato de Poliglicerila é um ingrediente vegetal com propriedades antimicrobianas utilizado como desodorante em formulações cosméticas. Ele atua combatendo a ação de bactérias causadoras de odor na pele em concentrações muito baixas (1 a 2 %) (GOLDSCHMIDT, 2003).

Segundo informações do fornecedor da matéria prima e pelo histórico da empresa, para o Caprilato de Poliglicerila não existe procedimento padrão de quantificação. Análises cromatográficas (HPLC) preliminares apontaram a presença de quatro picos adjacentes, indicando uma composição complexa. Este comportamento se deve ao fato do ingrediente ser uma mistura de moléculas de poligliceril em diferentes graus de condensação, com maior incidência de triglicerol, esterificadas com ácido caprílico (cadeia saturada de oito carbonos) (FIGURA 9).

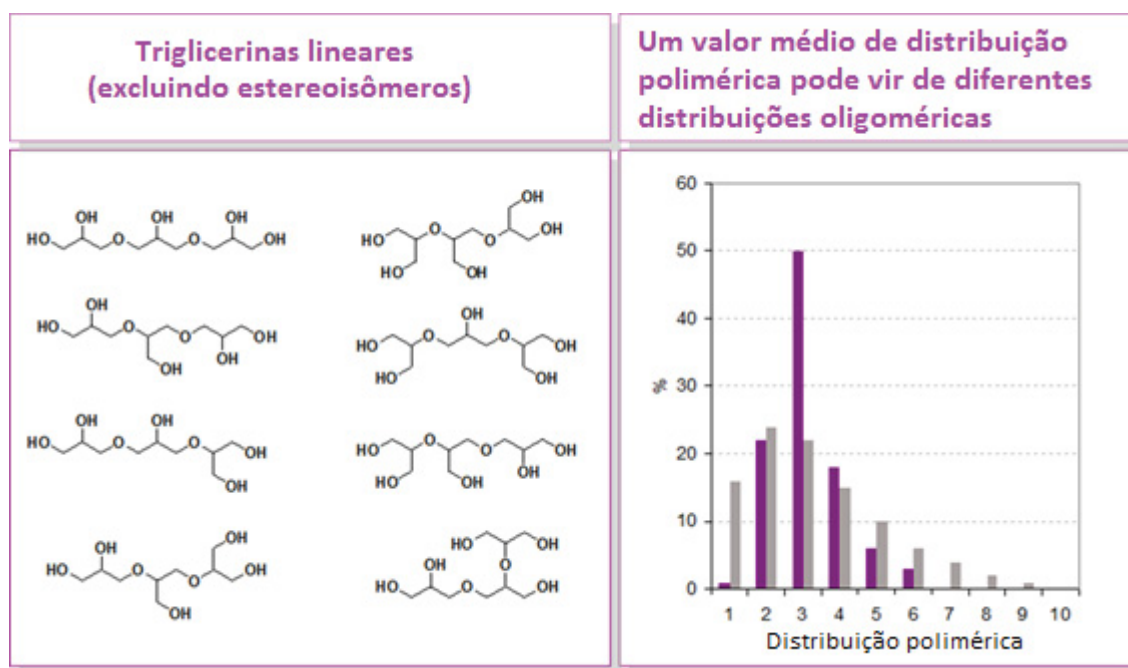


FIGURA 9 –INFORMAÇÕES SOBRE O CONSERVANTE CAPRILATO DE POLIGLICERILA

FONTE: (GOLDSCHMIDT, 2003)

5.1. MATERIAL E MÉTODOS

5.1.1. Programas computacionais

Para a montagem das matrizes de dados foi utilizado o software Origin Pro 9.1[®] (OriginLab), enquanto que para a elaboração dos modelos foi empregado o pacote PLS-Toolbox 3.0 (Eigenvector Research, Inc.) que opera em ambiente Matlab[®] v.7.0.1 (Math Work Inc.).

5.1.2. Reagentes e insumos

O Caprilato de Poliglicerila empregado nas análises, assim como, as demais matérias primas, foi fornecido pelo Grupo Boticário em grau de pureza adequado para uso cosmético.

Para as análises de quantificação do Caprilato de Poliglicerila foi utilizada uma composição de creme hidratante com os componentes comumente aplicados em composições cosméticas, conforme demonstrado na TABELA 4. Tal composição foi analisada pela aplicação direta de uma alíquota da mesma sobre o cristal do equipamento para coleta dos espectros de infravermelho.

TABELA 4 - COMPOSIÇÃO QUALITATIVA DO CREME HIDRATANTE EMPREGADO

Componente	Aplicação
Água	Solvente
Álcool cetílico	Emoliente / emulsionante
Ácido esteárico	Emoliente / emulsionante
BHT (hidroxitolueno butilado)	Antioxidante / mascarante
Vermelho ácido 33 ci17200	Corante
Trietanolamina	Ajuste de pH
Azul brilhante ci42090	Corante
Estearato de sorbitano	Emulsionante
Carbomer etd 2020	Estabilizante de emulsão /controlador de viscosidade
Aloe 200:1	Hidratante / anti-inflamatório
Edetato dissodico	Quelante / controlador de viscosidade
Dimeticona	Emoliente
Salicilato de octila	Filtro ultravioleta
Estearato de glicerila vegetal	Emoliente / emulsionante
Glicerol	Umectante
Amido-octenilsuccinato de alumínio	Controlador de viscosidade
Álcool benzílico	Conservante
Acrilat acril taurato de sodio	Estabilizante de emulsão /controlador de viscosidade
P-metoxicinamato de 2-etilexila	Filtro ultravioleta
Butilmetoxidibenzoilmetano	Filtro ultravioleta
Éter dicaprílico	Emoliente
Polisorbato-60	Emulsionante
Vaselina líquida usp 120	Emoliente
Fenoxietanol	Conservante
Extrato da flor de camomila recutita	Mascarante / condicionante
Essência de blueberry	Fragrância
Óleo de amêndoas	Emoliente

Sobre as porções da fórmula base do creme, foram adicionadas quantidades do conservante Caprilato de Poliglicerila de maneira a se obter concentrações massa/massa (mg/g) crescentes, sendo as mesmas então separadas em dois conjuntos: um de calibração contendo 11 amostras e um conjunto de teste contendo 4 amostras (TABELA 5).

TABELA 5 - CONCENTRAÇÃO DE CAPRILATO DE POLIGLICERILA NAS AMOSTRAS

Conjunto	Concentração (mg/g)
Treino	0.67
Treino	0.72
Treino	0.91
Treino	1.16
Treino	1.02
Treino	1.12
Treino	1.23
Treino	1.41
Treino	1.51
Treino	1.62
Treino	1.82
Teste	0.81
Teste	1.01
Teste	1.31
Teste	1.72

5.1.3. Desenvolvimentos de modelos de calibração multivariada

Para as análises envolvendo ATR, foi utilizado um espectrofotômetro de infravermelho com transformada de Fourier Spectrum 400 MID/NIR da PERKIN ELMER, equipado com acessório para medidas por reflectância total atenuada e cristal de diamante, com resolução de 2 cm^{-1} e 10 varreduras na região entre 4000 e 400 cm^{-1} .

Visando determinar os valores de concentração do conservante nas amostras através de uma análise PLS, vários tipos de pré-processamento de sinais (ex. alisamento, derivada, correção multiplicativa de sinal, etc.) foram testados e o número de variáveis latentes foi selecionado em função do menor erro de previsão (RMSECV) na etapa de validação cruzada.

5.2. RESULTADOS E DISCUSSÃO

5.2.1. Quantificação do Caprilato de Poliglicerila através da regressão por mínimos quadrados parciais (PLS)

Visando estabelecer um procedimento de quantificação do conservante Caprilato de Poliglicerila em composições cosméticas, foram realizados estudos iniciais visando definir as regiões do espectro de infravermelho que melhor refletiam as variações de concentração do mesmo. Na FIGURA 10 (A) são apresentados os espectros das 11 amostras do conjunto de calibração e na FIGURA 10 (B) a média dos espectros referentes às amostras é comparada ao espectro do Caprilato de Poliglicerila puro.

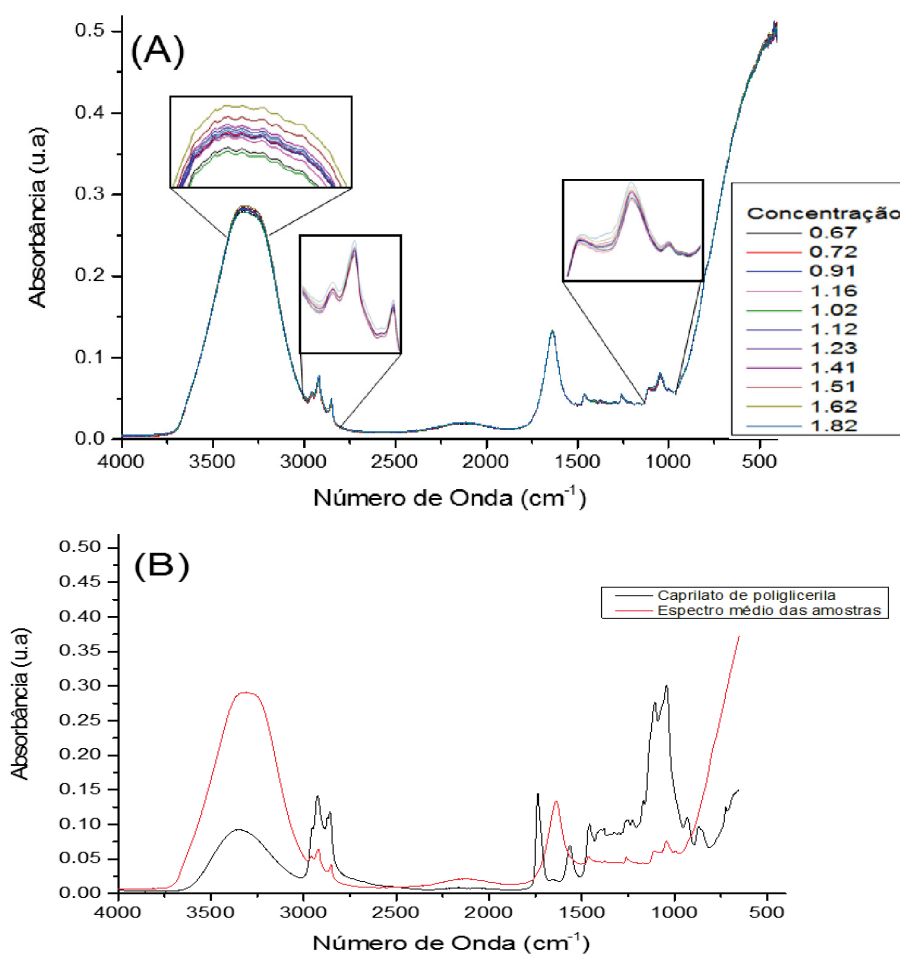


FIGURA 10 – (A) ESPECTROS DE INFRAVERMELHO DAS 11 AMOSTRAS DO CONJUNTO DE CALIBRAÇÃO. (B) COMPARAÇÃO ENTRE O ESPECTRO MÉDIO DAS AMOSTRAS E O ESPECTRO DO PADRÃO DO CONSERVANTE

Conforme observado na FIGURA 10 (A) e (B), as regiões do espectro com maior potencial de conter informação sobre variações na concentração do Caprilato de Poliglicerila seriam: 3350 cm^{-1} (estiramento O–H), o intervalo entre 2800 a 2900 cm^{-1} (estiramentos simétricos e assimétricos de ligações C–H), 1740 cm^{-1} (estiramento da C=O carbonílico) e a região entre 1110 e 1040 cm^{-1} (estiramento C–O), uma vez que nesses pontos observam-se os máximos de absorção para a molécula do conservante.

Após serem avaliadas diversas combinações de pré-tratamentos para os dados brutos, optou-se por realizar o menor número possível de transformações nos dados, visando a obtenção de um modelo de regressão simples e com menor risco de sobreajuste, procedendo-se então pré-processamento com dados centrados na média dos blocos X e Y.

De posse dos espectros coletados e pré-processados, empregou-se a regressão por PLS segundo o algoritmo de NIPALS, a qual se baseia em processos matemáticos de redução de dimensionalidade, rotação dos eixos e busca pela correlação entre a matriz de dados X (espectros processados) e a matriz de resultados Y, a qual contém os valores de concentração da espécie de interesse nas amostras do conjunto de treinamento.

Para auxiliar na escolha do número de variáveis latentes, foi empregado o procedimento de validação cruzada *leave one out*. Os erros que surgem desta previsão são apresentados como a raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) em função do número de variáveis latentes (FIGURA 11), sendo que duas LVs foram definidas como a combinação que mais reduziu esse valor de erro. Portanto, o modelo foi construído utilizando-se duas variáveis latentes, as quais explicam 97,24 % da informação contida na matriz X e 96,69 % da informação contida na matriz Y.

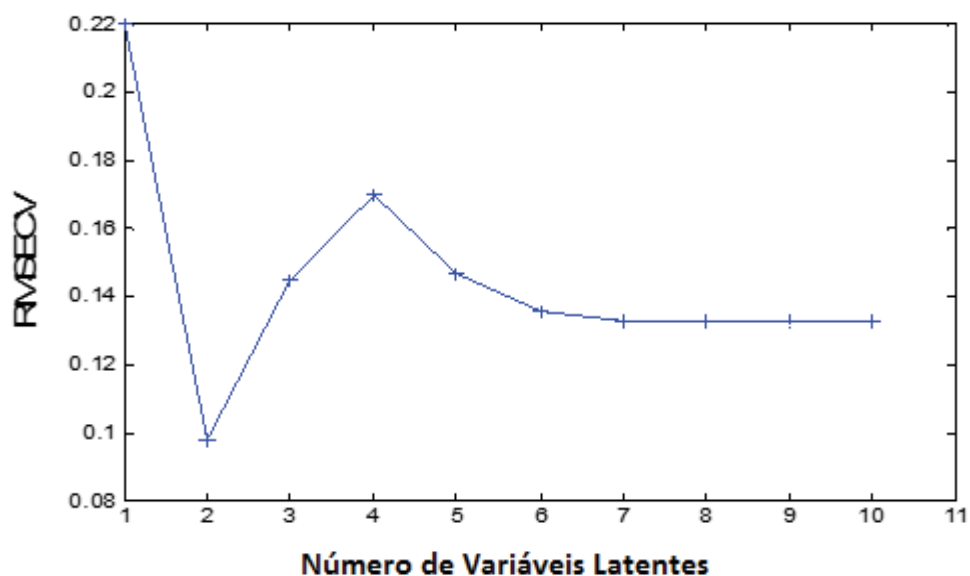


FIGURA 11 – RMSECV VS. NÚMERO DE VARIÁVEIS LATENTES PARA O MODELO

Na FIGURA 12 é representado o vetor de regressão do modelo para duas variáveis latentes.

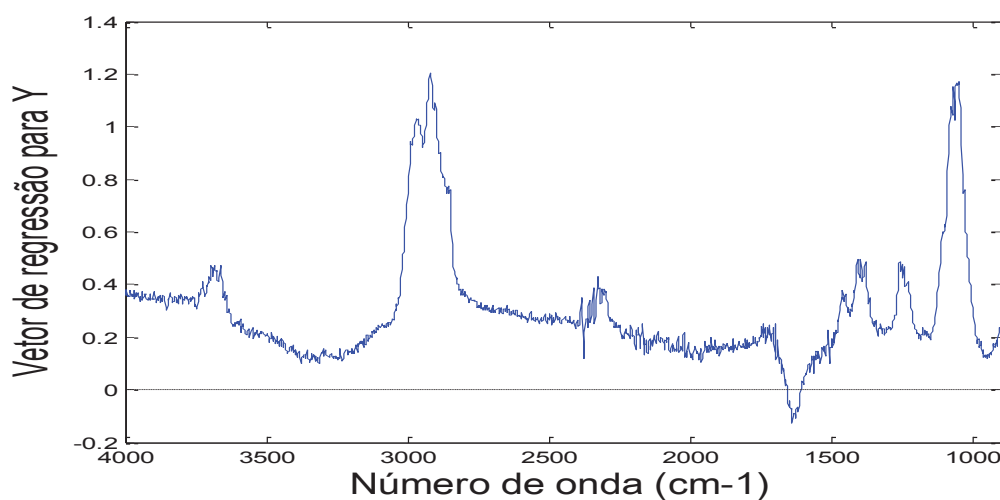


FIGURA 12 – VETOR DE REGRESSÃO DO MODELO DE PLS PARA QUANTIFICAÇÃO DO CAPRILATO DE POLIGLICERILA

Um ponto a ser destacado é o fato do vetor de coeficiente de regressão considerar como mais importantes para a correlação desejada, as regiões de absorção do Caprilato de Poliglicerila: 3700 cm^{-1} (início da banda com máximo em 3350 cm^{-1}), o intervalo entre 2800 a 2900 cm^{-1} e a região entre 1110 e 1040 cm^{-1} . Este antecedente evidencia que o modelo construído é condizente com o objetivo

proposto. Para detecção de amostras anômalas, o gráfico de *leverage* vs. resíduos studentizados foi analisado (FIGURA 13). Nenhuma amostra apresentou valor de resíduo elevado (maior que $\pm 2,5$). Quanto a *leverage*, o limite obtido pela EQUAÇÃO 6, foi de 0.54, indicando que uma amostra apresenta alta *leverage*. Como a *leverage* representa a distância da amostra ao centróide do conjunto de dados, amostras que contém a menor e/ou a maior concentração apresentam a tendência de ter uma *leverage* mais elevada. Tal fato foi analisado, e como a referida amostra apresenta a maior concentração do Caprilato de Poliglicerina, optou-se por não a eliminar para a construção do modelo.

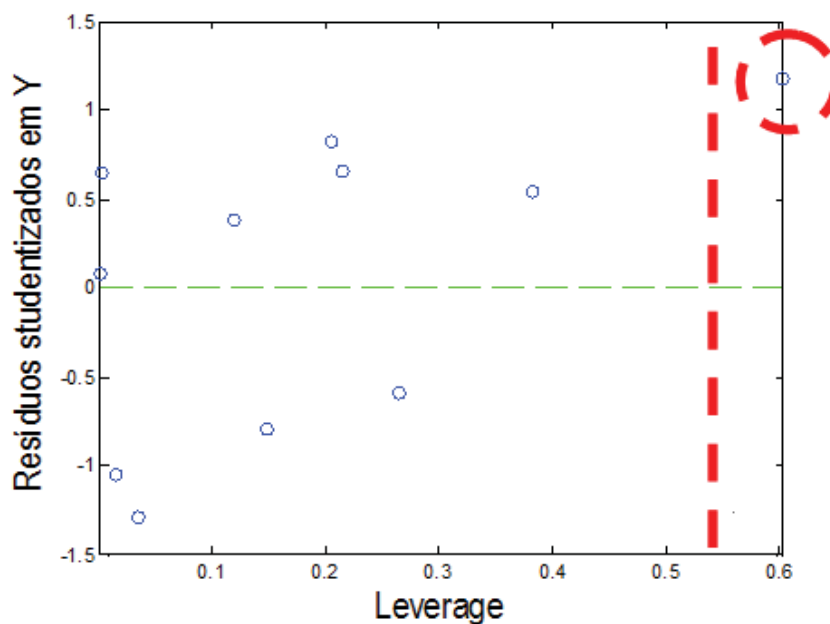


FIGURA 13 – REPRESENTAÇÃO GRÁFICA DA ANÁLISE DE RESÍDUOS STUDENTIZADOS VERSUS LEVERAGE

O modelo com as características descritas anteriormente apresentou a seguinte correlação entre os valores de concentração medidos contra os previstos (FIGURA 14) para as amostras de creme do conjunto de calibração via validação cruzada (*leave one out*).

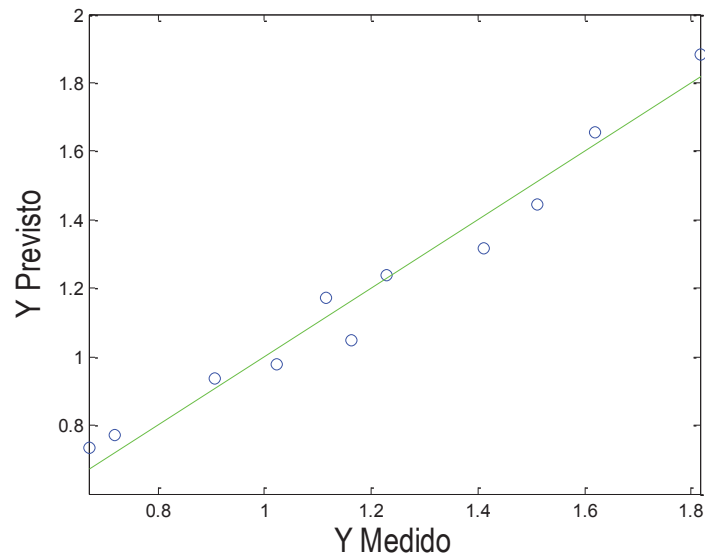


FIGURA 14 - COMPARAÇÃO ENTRE AS OS VALORES DE CONCENTRAÇÃO MEDIDOS CONTRA OS PREVISTOS PARA AS AMOSTRAS DE CREME. VALOR DE $R^2=0,968$

No procedimento de validação cruzada, o modelo PLS apresentou um valor de erro máximo de 9,6 % na previsão dos valores de concentração do Caprilato de Poliglicerila em amostras de creme hidratante. Este valor, embora maior do que os 5 % usualmente aceitos para métodos de quantificação, representa um resultado extremamente positivo, considerando-se que não há métodos viáveis para o controle de qualidade da matéria prima de interesse. Com base nas informações obtidas na modelagem via conjunto de calibração, o modelo construído foi então empregado para analisar as amostras do conjunto teste (quatro amostras cujos espectros não foram empregados na construção do modelo). A TABELA 6 apresenta os resultados da aplicação do modelo a este conjunto, com os valores de concentração previstos para o Caprilato de Poliglicerila e o erro relativo dos mesmos em relação aos valores reais.

TABELA 6 - VALORES REAIS E PREVISTOS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA

Real	Previsto	Erro relativo (%)
0,81	0,86	6,17
1,01	1,09	7,92
1,31	1,31	0,00
1,72	1,63	-5,23

Os resultados obtidos para estas amostras demonstraram que o modelo construído foi capaz de prever com boa precisão os valores de concentração do analito em amostras complexas de creme hidratante.

5.3. CONSIDERAÇÕES FINAIS E CONCLUSÕES

Foi desenvolvida uma metodologia quantitativa via PLS, a qual foi bem-sucedida na determinação dos valores de concentração para o conservante Caprilato de Poliglicerila em uma matriz cosmética referente à composição básica de um creme hidratante. A importância desse método se destaca devido à baixa concentração do analito e a dificuldade de análise da matéria prima via métodos cromatográficos, constituindo um desafio analítico para a indústria cosmética.

Capítulo III – SISTEMA RECOMENDADOR DE EMOLIENTES

6. SISTEMA RECOMENDADOR DE EMOLIENTES

Emolientes são componentes fundamentais nas formulações cosméticas, representando a segunda maior classe de componentes em concentração (geralmente de 3 a 20 % m/m), sendo utilizados apenas em menor quantidade que a água (SAVARY *et al.*, 2013). Contribuem para a hidratação, lubrificação, proteção, condicionamento e performance, além de serem elementos chaves para o sensorial, atributo muito importante para a aceitação dos produtos pelos consumidores (FITOUSSI, 2011).

Existe uma vasta gama de emolientes disponíveis no mercado, dentre eles os mais comuns são os ésteres, éteres, silicones, óleos, manteigas, hidrocarbonetos, parafínicos, derivados de lanolina, entre outros (SAVARY *et al.*, 2013). Porém, em sua maioria, faltam informações e dados técnicos extremamente importantes para a correta escolha do melhor emoliente para cada uso, visando a segurança, eficácia, estabilidade e sensoriais desejados. Nesse contexto, a construção de um sistema capaz de recomendar a melhor aplicação para cada emoliente com base nas suas propriedades sensoriais e físico-químicas aparece como uma oportunidade para a aplicação de ferramentas multivariadas de maneira a aproveitar ao máximo a informação fornecida pelas diferentes análises.

6.1. MATERIAL E MÉTODOS

6.1.1. Programas computacionais

Para a montagem das matrizes de dados, elaboração dos modelos, execução das análises e construção de gráficos foi empregado o software R (R CORE TEAM, 2015).

6.1.2. Reagentes e insumos

Os emolientes empregados nas análises foram fornecidos pelo Grupo Boticário em grau de pureza adequado para uso cosmético. Foram selecionados 38 emolientes cosméticos, os quais se encontram agrupados de acordo com a classificação química na TABELA 7.

Capítulo III – Sistema recomendador de emolientes

TABELA 7 - CLASSIFICAÇÃO QUÍMICA DOS EMOLIENTES SELECIONADOS

Classificação	Código da amostra	Emoliente
Ésteres	4	Adipato de Dibutila
	10	Adipato de diisopropila
	11	Benzoato de alquila C12-15
	12	Carbonato de dicaprila
	20	Dicaprilate Dicaprato neopentil glicol
	21	Estearato de isocetilestearyl
	13	Estearato de octila
	31	Isononanoato de cetostearyl
	14	Isononanoato de octila
	38	Malato de diisostearyl
	15	Miristato de isopropila
	16	Neopentanoato de isodecila
	22	Oleato decila
Álcool	19	Álcool octildodecílico
Éteres	1	Éter dicaprílico
Silicones	3	Ciclometicone
	29	Dimeticona
	30	Feniltrimeticone
Óleos e Glicerídeos	8	Cocoglicerídeos
	9	Esqualano vegetal
	24	Óleo amêndoas doces
	25	Óleo de abacate
	26	Óleo de girassol
	27	Óleo de jojoba
	37	Óleo de rícino
	28	Óleo de semente de uva
	35	Triglicerídeos cáprico-caprílico
Hidrocarbonetos e parafínicos	5	Isohexadecano
	33	Vaselina líquida USP 120
	34	Vaselina líquida USP 80/90
Derivados de Lanolina	7	Álcool de lanolina acetilado
Ésteres Verdes	23	Beantree (Methylheptyl isostearate)
	17	Dermofeel Sensolv (Isoamyl Laurate)
	18	Dub Zenoat Propanediol dicaprylate
	32	Schercemol 1818 (Isostearyl isostearate)
	36	Schercemol CATC (Cocoyl Adipic Acid/Trimethylolpropane Copolymer)
	2	Schercemol PTID (Triisostearyl Polyglyceryl-3-dimer dilinoleate)
	6	Tegosoft AC (Isoamyl cocoate)

Capítulo III – Sistema recomendador de emolientes

6.1.3. Análises sensoriais e físico-químicas

Com o objetivo de definir quais os usos mais indicados para cada emoliente (maquiagem, cremes, proteção solar, etc.), foi realizada uma série de análises de acordo com os protocolos internos da empresa. Na TABELA 8 são apresentadas as técnicas empregadas para as medidas realizadas, assim como as informações fornecidas por cada uma.

TABELA 8 - ANÁLISES REALIZADAS PARA A CARACTERIZAÇÃO DOS EMOLIENTES

Análise	Metodologia	Objetivo	Aplicações prática
Índice de refração	Refratometria	Caracterizar os emolientes quanto ao alto ou baixo índice de refração	<ul style="list-style-type: none"> • Efeitos: • Brilho • Matte
Espalhabilidade	Ângulo de contato	Caracterizar os emolientes quanto à baixa, média ou alta espalhabilidade	<ul style="list-style-type: none"> • Sensorial: • Leve • Médio • Pesado
Tensão interfacial	Curvatura da gota	Caracterizar os emolientes quanto à baixa ou alta tensão interfacial e quão fáceis eles são de emulsionar	<ul style="list-style-type: none"> • Estabilidade • Facilidade em emulsionar
Polaridade	Curvatura da gota	Caracterizar os emolientes quanto à alta, média ou baixa polaridade (quais são os mais apolares e quais são os mais polares)	<ul style="list-style-type: none"> • Estabilidade de fórmulas; • Solubilização e compatibilidade com filtros solares, pigmentos, etc.

Adicionalmente, os valores de densidade e viscosidade para cada matéria prima foram coletados.

6.2. RESULTADOS E DISCUSSÃO

6.2.1. Pré-tratamento dos dados para análise de componentes principais

Devido à disparidade dos valores de viscosidade (2,67 a 17809,00 cP), sem uma distribuição regular, uma escala em níveis foi criada com valores de: 1 para a faixa de 0 a 10,00 cP; 2 para valores entre 10,01 e 100 cP; 3 para valores entre 100,01 e 500,00 cP; 4 para valores entre 500,01 e 1000,00; e 5 para valores acima de 1000,00 cP, conforme demonstrado na TABELA 9.

Após serem avaliadas diversas combinações de pré-tratamentos para os dados, a PCA foi realizada, sendo que a opção que gerou os melhores resultados na discriminação dos grupos foi o autoescalamento.

Capítulo III – Sistema recomendador de emolientes

TABELA 9 – RESULTADOS DAS ANÁLISES DOS EMOLIENTES

Código da amostra	Densidade (g/cm³)	Índice de refração	Ângulo de contato (°)	Tensão interfacial	Viscosidade (cP)	Viscosidade *
1	0,807	1,4230	13,3	72,35	3,87	1
2	0,960	1,4736	66,1	3,41	17809,00	5
3	0,955	1,3963	0	33,82	3,94	1
4	0,960	1,4340	7,5	2,32	2,67	1
5	0,790	1,4392	8,8	2,66	4,55	1
6	0,850	1,4373	11,7	9,98	5,95	1
7	0,869	1,4417	9,2	11,54	8,24	1
8	0,935	1,4523	17,8	7,79	35,90	2
9	0,813	1,4571	18,4	11,20	36,70	2
10	0,959	1,4318	11,4	19,96	4,22	1
11	0,925	1,4833	14,2	24,82	11,40	2
12	0,891	1,4345	10,4	29,87	6,09	1
13	0,860	1,4470	12,3	31,61	11,90	2
14	0,855	1,4346	10,6	31,68	3,70	1
15	0,852	1,4322	9,6	25,80	5,39	1
16	0,853	1,4283	12,9	28,30	4,17	1
17	0,880	1,4348	11,9	22,96	6,84	1
18	0,930	1,4407	12,0	23,33	9,94	1
19	0,840	1,4524	17,7	25,57	53,20	2
20	0,920	1,4425	20,9	26,29	11,10	2
21	0,875	1,4580	23,1	28,38	73,60	2
22	0,865	1,4555	17,6	29,86	14,00	2
23	0,855	1,4471	21,1	30,85	15,70	2
24	0,915	1,4706	18,6	26,15	52,30	2
25	0,915	1,4690	21,7	23,35	58,10	2
26	0,905	1,4681	18,0	22,70	71,50	2
27	0,885	1,4645	17,6	22,89	31,40	2
28	0,916	1,4726	24,8	25,30	43,60	2
29	0,969	1,4036	17,6	34,23	317,70	3
30	0,980	1,4594	18,9	28,54	19,50	2
31	0,854	1,4463	16,3	16,76	16,10	2
32	0,865	1,4594	21,3	17,95	41,20	2
33	0,852	1,4716	21,5	18,91	36,60	2
34	0,849	1,4661	15,9	17,10	21,30	2
35	0,945	1,4475	18,9	17,40	21,30	2
36	0,950	1,4616	25,5	12,37	234,30	3
37	0,963	1,4778	29,1	16,76	549,30	4
38	0,915	1,4591	52,2	24,13	2321,00	5

6.2.1. Análise Hierárquica de Agrupamentos (HCA)

Conforme observado na FIGURA 15, o dendrograma mostra a similaridade entre os emolientes avaliados, sendo que a escala é dada pela distância entre os agrupamentos (0 a 5), com o valor máximo de similaridade dado pela distância igual a zero. A maior similaridade obtida foi entre os emolientes em quatro grandes agrupamentos (clusters 7, 9, 10 e 11), com uma distância padrão de 2 (60 % de similaridade), pois esta foi a que melhor separou esses agrupamentos de acordo com a similaridade. As demais amostras foram classificadas em agrupamentos individuais.

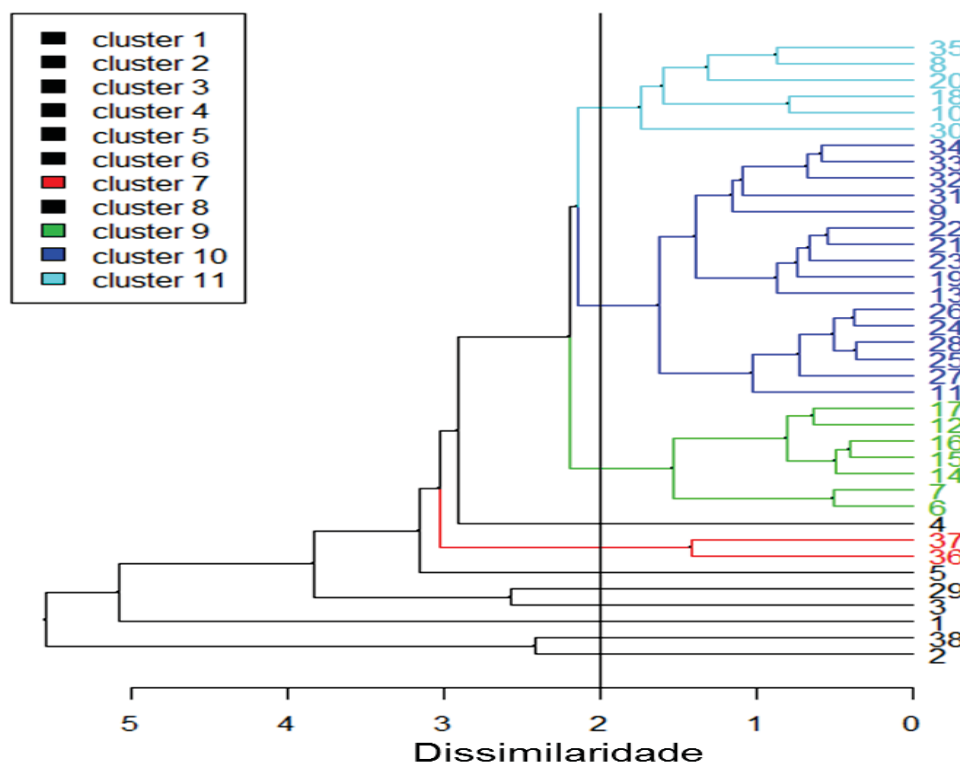


FIGURA 15 – DENDROGRAMA OBTIDO PARA AS 38 AMOSTRAS DE EMOLIENTES. A ESCALA ABAIXO DO DENDROGRAMA INDICA OS ÍNDICES DE DISTÂNCIA ENTRE AS AMOSTRAS (SIMILARIDADE MÁXIMA PARA DISTÂNCIA IGUAL A ZERO)

Os resultados gerados pela HCA podem ser empregados para indicar os melhores substitutos para cada matéria prima, por exemplo, os emolientes 33 e 34 foram apontados como muito semelhantes. Ao se analisar diretamente seus dados físico-químicos, constata-se que essas amostras são dois tipos de vaselina líquida que diferem entre si apenas levemente pelos valores de viscosidade e ângulo de contato.

A visualização de similaridades de maneira estruturada oferecida pela HCA não oferece explicações de quais variáveis são responsáveis pelas separações entre grupos. Nesse contexto, a PCA além de confirmar a validade dos agrupamentos gerados pela HCA devido ao fato de chegarem às mesmas distribuições baseando-se em operações matemáticas diferentes (a saber, cálculo de distância euclidiana e redução de dimensionalidade), surge como análise complementar, em especial pelas informações que a matriz de *loadings* traz sobre a relação entre a separação dos os grupos e quais variáveis originais são responsáveis por tais diferenciações.

6.2.2. Análise por componentes principais

As duas primeiras componentes principais (PCs) explicam 70 % da variância do sistema. Em outras palavras, 70 % da informação presente nas cinco medidas originais tem quase toda a sua informação explicada por apenas duas dimensões, as quais podem ser visualizadas em um gráfico bidimensional.

A FIGURA 16 contém a apresentação gráfica dos *scores* das componentes principais (PCs) 1 e 2 geradas pela PCA. Os mesmos agrupamentos observados na HCA podem ser visualizados, com os clusters 7, 9, 10 e 11 se destacando das amostras com valores de *scores* mais dispersos para as PCs 1 e 2 (emolientes 1, 2, 3, 5, 29 e 30).

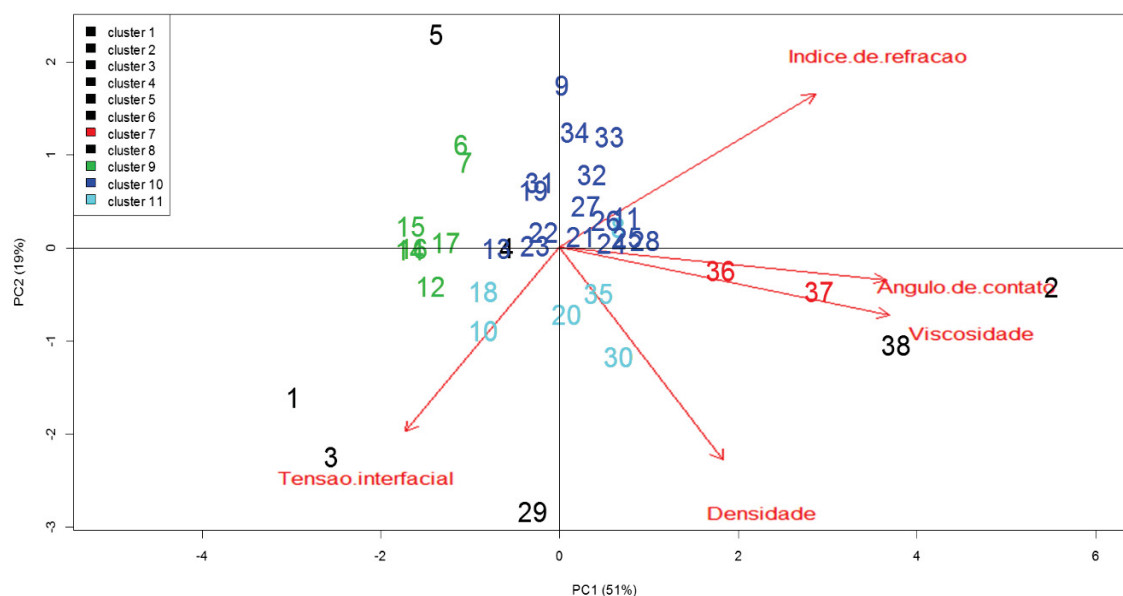


FIGURA 16 – APRESENTAÇÃO GRÁFICA DOS SCORES (NÚMEROS EM CORES) E LOADINGS (FLECHAS EM VERMELHO) PARA AS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS

Visando entender os aspectos físico-químicos que diferenciam os grupos, uma análise dos gráficos dos *loadings* das respectivas PCs foi realizada. Como os *loadings* representam o peso que cada variável original tem na construção das PCs, uma análise desses dados evidencia quais variáveis originais estão sendo consideradas importantes para a separação dos grupos obtidas no gráfico de *scores*.

Conforme demonstrado na FIGURA 16, a PC1 considera relevante a informação de todas as variáveis originais, mas leva em conta majoritariamente a informação contida nos ensaios de ângulo de contato e viscosidade, as quais fornecem informação principalmente sobre a espalhabilidade dos emolientes (WORTEL; WIECHERS, 2000). Valores elevados de ângulo de contato refletem baixos valores de espalhabilidade, levando a um sensorial mais pesado. A tendência a um sensorial mais leve foi a diferença mais significativa entre as amostras do cluster 9 e os demais. Um exemplo que confirma este fato é que as amostras no quadrante negativo da PC1 são as que apresentam os menores valores de ângulo de contato, como a amostra 3 (ciclometicone), a qual apresenta o menor ângulo de contato entre todas as amostras. Enquanto que as amostras com os maiores ângulos de contato – 2 (Schercemol PTID) e 38 (Malato de diisoestearila) – ficaram no extremo direito do quadrante positivo da PC1 e apresentam os menores valores de espalhabilidade e consequentemente levando a um sensorial mais pesado.

A PC2 levou em conta majoritariamente a informação das medidas do índice de refração, densidade e tensão interfacial, sendo esta última a que melhor explicou a separação entre as amostras por estar diretamente relacionada com a polaridade das substâncias envolvidas, sendo que a estabilidade das fórmulas produzidas está diretamente relacionada à polaridade dos emulsionantes empregados (USHIKUBO; CUNHA, 2014), conforme descrito na TABELA 10. A PC2 levou à separação das amostras dos clusters 10 e 11 e à uma distribuição de todas as amostras ao longo do eixo da componente, sendo que a amostra 5 (Isohexadecano) apresenta um elevado índice de refração, baixa tensão interfacial e um baixo valor de densidade, como esperado da amostra com maior *score* na PC2. Da mesma maneira, a amostra 29 (dimeticona), que possui o valor mais negativo de *score* na PC2, apresenta tendências de valores opostos aos das medidas da amostra 5.

Capítulo III – Sistema recomendador de emolientes

Para aumentar a aplicabilidade da metodologia proposta, a TABELA 10 fornece informações obtidas experimentalmente e baseadas na experiência de formuladores do Grupo Boticário.

TABELA 10 – RELAÇÕES OBTIDAS

Análise	Aplicações prática	Relação
Índice de refração	Efeitos: ✓ Brilho ✓ Matte	↑ IR ↑ Brilho ↓ Matte ↓ IR ↓ Brilho ↑ Matte
Espalhabilidade	Sensorial: ✓ Leve ✓ Médio ✓ Pesado	↑ θ ↓ espalhabilidade (sensorial mais pesado) ↓ θ ↑ espalhabilidade (sensorial mais leve)
Tensão interfacial	✓ Estabilidade ✓ Facilidade em emulsionar	↑ TI ↓ facilidade de emulsionar ↓ estabilidade ↓ TI ↑ facilidade de emulsionar ↑ estabilidade
Polaridade	✓ Estabilidade de fórmulas; ✓ Solubilização e compatibilidade com filtros solares, pigmentos, etc.	↑ TI ↑ caráter apolar ↓ TI ↑ caráter polar

A metodologia proposta para a substituição de uma matéria prima existente ou para recomendação de aplicação para uma nova substância é iniciada pela realização das análises mencionadas; processamento via HCA e PCA. Primeiramente, é feita a identificação dos emolientes mais semelhantes ao de interesse (HCA). Em seguida, uma análise dos *loadings* da PCA fornece informação a respeito de quais as diferenças mais significativas em termos de propriedades físico-químicas e a TABELA 10 aponta quais as diferenças em termos de propriedades podem ser esperadas.

6.3. CONSIDERAÇÕES FINAIS E CONCLUSÕES

Devido à grande importância dos emolientes para a indústria cosmética e com o objetivo de agrupar os mesmos de acordo com a sua aplicação em diferentes classes de produtos, uma análise exploratória de dados físico-químicos foi realizada baseando-se no algoritmo HCA que apresentou resultados positivos no processo de agrupamento. A ferramenta se mostrou valiosa no desenvolvimento de novas fórmulas e estudo de substituição e contratipagem de matérias primas. Conhecer individualmente e profundamente os componentes de uma formulação cosmética contribui imensamente para o sucesso do produto final.

Os resultados dos testes realizados permitiram ranquear os 38 emolientes selecionados com informações sobre índice de refração, viscosidade, espalhabilidade, sensorial e tensão interfacial, podendo ser usado como um guia para identificar o emoliente ideal para cada aplicação, assim como selecionar qual o melhor substituto para cada matéria prima. A ferramenta tornou possível entregar informações robustas sobre as características físico-químicas dos emolientes, sendo isso extremamente útil no dia-a-dia dos formuladores, aumentando a assertividade no desenvolvimento de formulações.

Capítulo IV – SABONETES EM BARRA

7. SABONETES EM BARRA

Sabonetes são sabões utilizados na higienização do corpo humano, podendo proporcionar benefícios adicionais, como hidratação e perfume. Do ponto de vista químico, o sabão é um sal de ácido graxo, obtido através do processo de saponificação. A saponificação é a reação entre uma gordura ou óleo e uma base (tradicionalmente hidróxido ou carbonato de sódio). Após a hidrólise dos precursores, os gliceróis livres ligam-se aos grupos hidroxila para formar glicerina, deixando em solução o sal graxo correspondente (sabão) (TECPAR, 2011).

A massa base para sabonetes tem três fontes principais: gordura animal, óleos vegetais e material sintético, sendo que este último é baseado em ésteres de isetionato, cujas características de pH neutro, facilidade de manipulação no ambiente fabril e sensorial agradável, o tornam uma opção interessante para confecção de sabonetes. Contudo, a massa sintética é mais cara e seu uso só se torna economicamente viável no caso de produtos artesanais de alto valor agregado ou produções em grande escala, como no caso do Grupo Unilever (JAMES *et al.*, 2000). As diferentes composições de óleos e gorduras produzem sabões com propriedades distintas, conforme apresentado na TABELA 11. Em geral, emprega-se uma combinação dessas matérias primas de forma a aliar as propriedades características de cada uma.

TABELA 11 - CARACTERÍSTICAS DOS ÓLEOS E GORDURAS GERALMENTE EMPREGADOS EM SABÕES

Gordura / Óleo	Tipo de espuma	Propriedades de limpeza	Ação sobre a pele
Sebo ou Sebo	Razoavelmente lenta, duradoura e espessa	Boa	Muito moderada
Coco	Espuma rapidamente com muitas bolhas não persistentes	Excelente	Ação mordente enrugando a pele
Palma	Espuma lentamente, bolhas pequenas e duradouras	Muito boa	Muito moderada
Babaçu	Espuma consistente de bolhas largas, não persistentes	Excelente	Ação mordente enrugando a pele
Algodão	Oleosa, abundante e durabilidade média	Boa	Moderada
Soja	Oleosa, abundante e duradoura	Regular	Moderada
Oliva	Gordurosa, bolhas pequenas e persistentes	Regular para boa	Muito moderada
Mamona	Espessa e duradoura	Regular	Moderada

Fonte: (MERCADANTE; ASSUMPÇÃO, 2010)

O fato das composições de sabonete em barra serem constituídas de misturas de óleos e gorduras saponificados, incluindo muitas vezes misturas de materiais proveniente de fontes animais, vegetais e/ou sintéticas faz deste um sistema de elevada complexidade para análises químicas. A composição dessas matérias primas varia não só em função do tamanho das cadeias, mas também no grau de instauração dos ácidos graxos que as compõem (TABELA 12), levando a produtos com características sensoriais bem diferentes. Embora os constituintes majoritários do sebo bovino possam se assemelhar aos de alguns óleos vegetais, há alguns componentes minoritários, tais como o colesterol, cuja presença em quantidades significativas só é esperada em massas de origem animal (O'BRIEN, 2009).

TABELA 12 - PRINCIPAIS DIFERENÇAS COMPOSICIONAIS ENTRE OS ÓLEOS E GORDURAS MAIS EMPREGADOS NA FABRICAÇÃO DE SABONETES

Teor dos ácidos graxos / análises de referência	Sebo (%)	Óleo de palma (%)	Óleo de coco (%)
C8:0 Caprilico	-	-	7.8
C10:0 Cáprico	-	-	6.7
C-12:0 Láurico	0.2	0	47.5
C-14:0 Mirístico	4	1.1	18.1
C-16:0 Palmítico	24.3	44	8.8
C-16:1 Palmitoleico	2.5	0.1	-
C-18:0 Esteárico	21.4	4.5	2.6
C-18:1 Oleico	33.6	39.2	6.2
C-18:2 Linoleico	1.6	10.1	1.6
C-18:3 Linolênico	0.6	0.4	
C-20:0 Araquídico	0.2	0.4	0.1
Valor de iodo	45	53	10
Colesterol (ppm)	1100	16	15

Fonte: (O'BRIEN, 2009)

Devido à natureza complexa da composição dos sabonetes (tanto a massa animal quanto a sintética contém uma porcentagem de massa vegetal), e ao fato de se apresentarem na forma sólida, análises exploratórias tradicionais via cromatografia líquida ou gasosa apresentam limitações devido à necessidade de etapas prévias de esterificação e extração.

7.1. MATERIAL E MÉTODOS

7.1.1. Equipamentos

Para as análises envolvendo ATR, foi utilizado um espectrofotômetro de infravermelho com transformada de Fourier Spectrum 400 MID/NIR da PERKIN ELMER, equipado com acessório para medidas por reflectância total atenuada e cristal de diamante, com resolução de 2 cm^{-1} e 10 varreduras na região entre 4000 e 400 cm^{-1} .

7.1.2. Programas computacionais

Para a montagem das matrizes de dados e construção dos gráficos foi utilizado o software Origin Pro 9.1[®] (OriginLab), enquanto que para a elaboração dos modelos foi empregado o pacote PLS-Toolbox 3.0 (Eigenvector Research, Inc.) que opera em ambiente Matlab[®] v.7.0.1 (Math Work Inc.).

7.1.3. Reagentes e insumos

As amostras de sabonete necessárias para a elaboração de modelos multivariados de classificação foram proporcionadas pelo Grupo Boticário ou adquiridos em mercados de Curitiba e região metropolitana.

Para o desenvolvimento dos modelos multivariados, foram adquiridas 74 amostras de sabonetes em barra, cujo critério e seleção foram baseados na disponibilidade dos mercados locais e a distribuição aleatória, visando cobrir as três principais classes de matéria prima (animal, vegetal e sintética).

7.1.4. Classificação das amostras e desenvolvimentos de modelos de calibração multivariada

As amostras de sabonete comercial foram classificadas de acordo com a natureza predominante da matéria prima (contendo massa base animal, vegetal ou sintética) com base na análise de rotulagem.

Para a construção dos modelos multivariados, foram elaborados conjuntos de calibração, contendo 16 amostras de sabonetes de massa base sintética, 22 amostras com massa vegetal e 36 amostras contendo massa base animal.

Estas amostras foram caracterizadas pela técnica espectroscópica de ATR. Os espectros foram processados integralmente.

Posteriormente, uma análise exploratória dos dados espectrais foi realizada, recorrendo-se a uma rotina de análise de PCA cujos resultados foram confrontados com as informações referentes à natureza da massa base usada na composição dos sabonetes. Vários tipos de pré-processamento de sinais (ex. alisamento, derivada, correção de sinal multiplicativa, etc.) foram testados visando a combinação que melhor ressaltasse as diferenças entre os espectros. Em seguida elaboraram-se modelos de classificação empregando-se a Análise Discriminante via Mínimos Quadrados Parciais (PLS-DA) e Modelagem Independente por Analogia de Classe (SIMCA). Ambos os modelos foram otimizados, levando-se em consideração suas particularidades e aplicados para classificação de amostras de sabonetes comerciais.

7.2. RESULTADOS E DISCUSSÃO

Estudos iniciais foram realizados para definir se os diferentes tipos de sabonetes poderiam ser diferenciados pelos respectivos espectros de infravermelho. Na FIGURA 17 são apresentados espectros representativos das três classes de sabonetes.

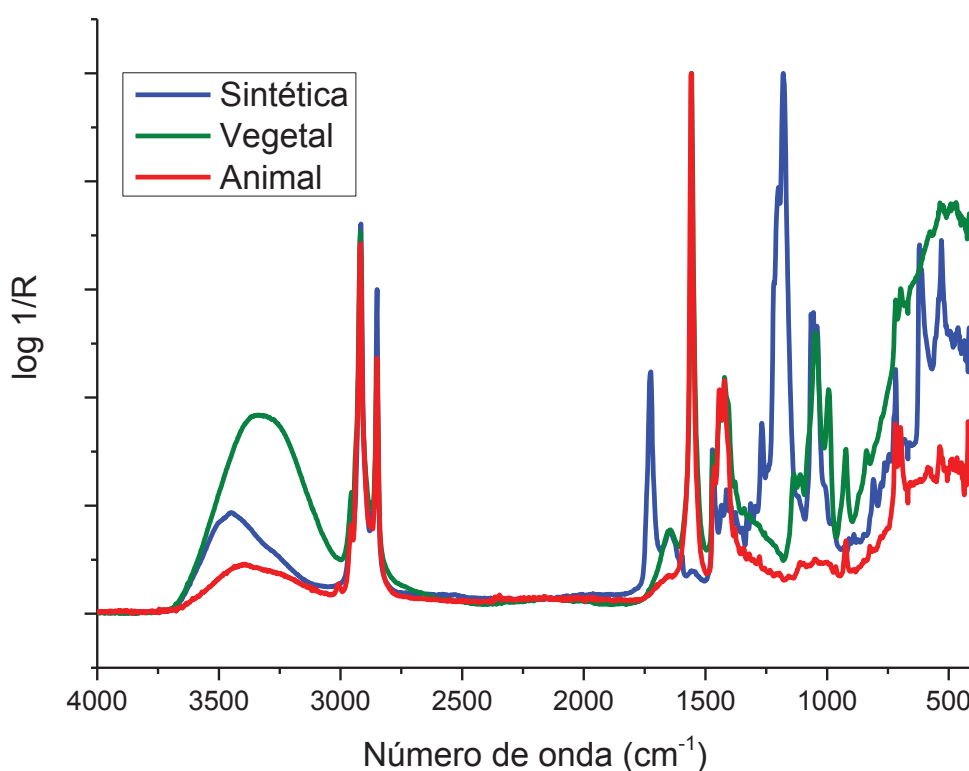


FIGURA 17 – ESPECTROS DE FTIR REPRESENTATIVOS DOS DIFERENTES GRUPOS DE SABONETES

7.2.1. Pré-tratamento dos dados para análise de componentes principais

Após serem avaliadas diversas combinações de pré-tratamentos para os dados brutos, a opção que gerou os melhores resultados na discriminação dos grupos foi centrar os dados na média.

Uma análise mais detalhada das diferentes regiões dos espectros de FTIR possibilitou destacar as diferenças presentes nos espectros de infravermelho de sabonetes contendo predominantemente massa base animal, vegetal e sintética,

Capítulo IV – Sabonete em barra

apontando a região entre 1850 e 800 cm^{-1} como a mais adequada para construção dos modelos, conforme demonstrado na FIGURA 18.

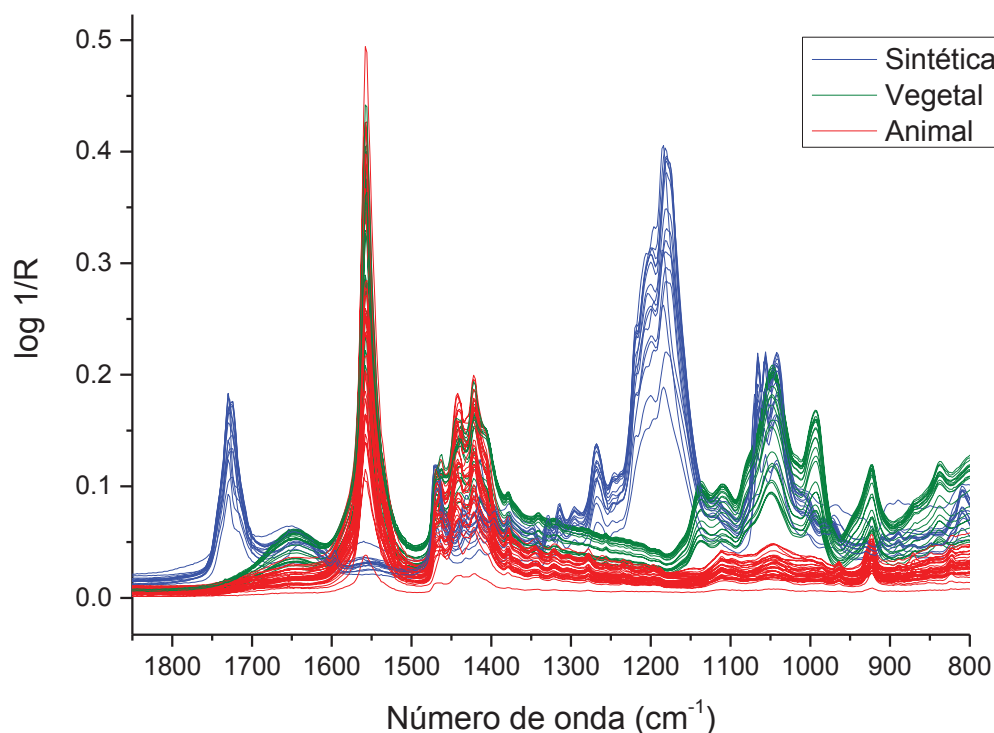


FIGURA 18 - REGIÃO DESTACADA DO ESPECTRO DE FTIR DAS AMOSTRAS

7.2.2. Análise por componentes principais

As duas primeiras componentes principais (PCs) explicam 95,09 % da variância do sistema. Em outras palavras, a porção do espectro de infravermelho situado entre 1850 e 800 cm^{-1} , antes representado por 526 pontos ou dimensões, depois de passar por uma série de transformações matemáticas, tem quase toda a sua informação explicada por apenas duas dimensões, as quais podem ser visualizadas em um gráfico bidimensional.

A FIGURA 19 contém a apresentação gráfica dos scores das componentes principais (PCs) 1 e 2 geradas pela PCA, conforme descrito. Podem ser observados três grupos divididos espacialmente, cada um deles representando uma das três classes de sabonete separados de acordo com a origem da matéria prima.

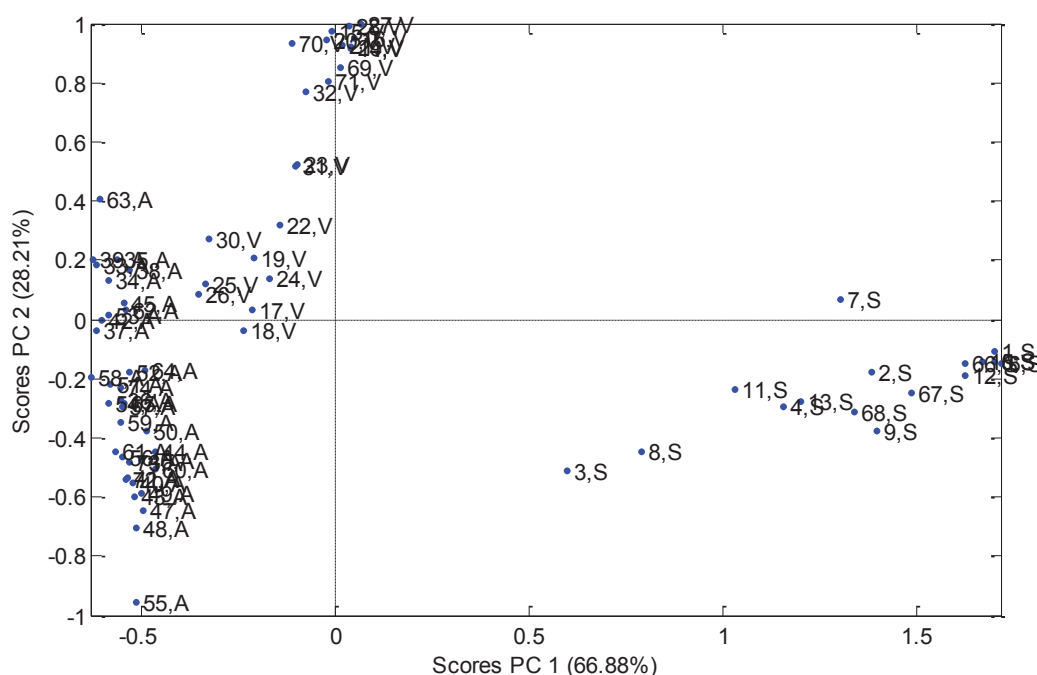


FIGURA 19 – APRESENTAÇÃO GRÁFICA DAS TRÊS PRIMEIRAS COMPONENTES PRINCIPAIS. AMOSTRAS PROVENIENTES DE SABONETES CONTENDO MATÉRIA PRIMA (A) ANIMAL; (V) VEGETAL; E (S) SINTÉTICA

Visando entender os aspectos químicos que diferenciam os grupos, uma análise dos gráficos dos *loadings* das respectivas PCs foi realizada. Como os *loadings* representam o peso que cada variável original (número de onda) tem na construção das PCs, uma análise desses dados evidencia quais regiões do espectro estão sendo consideradas importantes para a separação dos grupos obtidas no gráfico de *scores*.

Conforme demonstrado na FIGURA 20, a PC1 leva em conta majoritariamente as bandas em 1730, 1556, 1465, 1180, 1050 e 996 cm^{-1} . As bandas em 1730 e 1180 cm^{-1} foram atribuídas ao grupamento sulfonato característicos dos sais de isetionato (exclusivos da massa sintética) (PUTTNAM *et al.*, 1965). A banda em 1556 cm^{-1} foi atribuída aos grupos carboxílicos ionizados (sabão) (MIRGHANI *et al.*, 2002), o qual tende a se apresentar em maiores concentrações nas amostras contendo massa de origem vegetal e animal, uma vez que nas amostras contendo material sintético o sabão é substituído pelos ésteres de isetionato. O conjunto de bandas com máximo em 1465 cm^{-1} foi atribuído a deformações angulares dos grupamentos $-\text{CH}_2-$ das cadeias carbônicas

Capítulo IV – Sabonete em barra

(ROHMAN; CHE MAN, 2011). A banda em 1050 cm^{-1} foi atribuída aos estiramentos da ligação C-O devido a presença da glicerina (PUTTNAM *et al.*, 1965), a qual tende a estar presente em maior concentração nas amostras contendo material exclusivamente vegetal. Por fim, a banda em 996 cm^{-1} foi atribuída a deformações angulares das ligações C-H de cadeias carbônicas (SKOOG *et al.*, 2008), refletindo as diferenças composicionais das massas. Embora os componentes majoritários entre as massas sejam os mesmos, essa diferença de intensidade das bandas logo abaixo de 1000 cm^{-1} pode ser explicada por componentes minoritários como o colesterol, o qual está presente apenas em massas de origem animal (O'BRIEN, 2009).

A PC1 levou à separação das amostras contendo a massa base sintética das demais, assim como, há um princípio de separação entre as classes animal e vegetal. Enquanto a maioria das amostras de base vegetal apresenta *scores* positivos na PC2, a maior parte das amostras de base animal se concentra com *scores* negativos nesta PC.

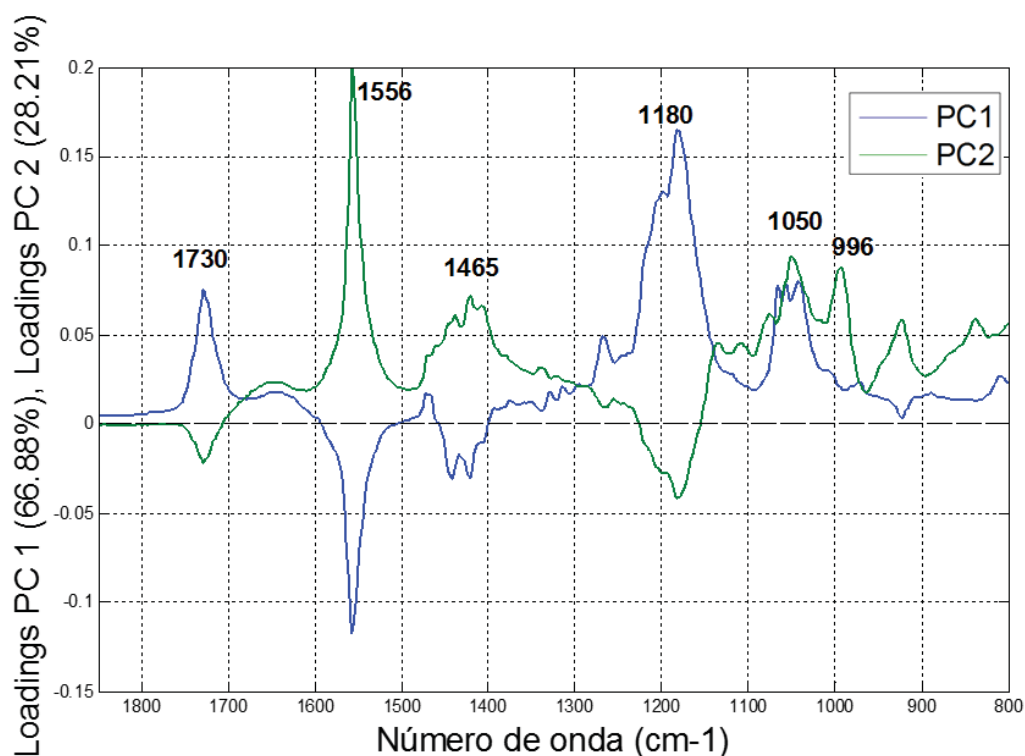


FIGURA 20 - LOADINGS DAS DUAS PRIMEIRAS COMPONENTES PRINCIPAIS (PCS 1 E 2)

7.2.3. Análise discriminante via mínimos quadrados parciais (PLS-DA)

De posse das informações obtidas através da análise de componentes principais e objetivando-se a classificação via espectros de ATR das amostras de sabonetes de interesse nos respectivos grupos (composição da matéria prima), foi empregado o método de PLS-DA, o qual se baseia nos mesmos processos matemáticos de redução de dimensionalidade que subsidiam a PCA, com a adição de um passo extra de rotação dos eixos e uma etapa de busca pela correlação entre a matriz de dados X (*scores* dos espectros processados) e os *scores* da matriz de resultados Y, a qual contém as classes das quais as amostras do conjunto de treinamento pertencem.

Para construção do modelo, o conjunto de dados experimentais foi dividido de maneira aleatória em dois: o primeiro, contendo 65 amostras empregadas para a etapa de calibração e um segundo, contendo 9 amostras (3 de cada grupo) para a etapa de previsão e verificação da capacidade de classificação do modelo.

Ao conjunto de calibração, foram empregados os mesmos pré-processamentos otimizados para a PCA, no caso: aproveitamento apenas da região do espectro entre 1850 e 800 cm^{-1} e centrar os dados na média. O bloco Y foi construído a partir da definição de classes: sabonete contendo matéria prima de origem sintética (Classe 1), sabonete contendo matéria prima de origem vegetal (Classe 2) e sabonete contendo matéria prima de origem animal (Classe 3). Esta matriz Y foi centrada na média e o algoritmo utilizado para a regressão foi o NIPALS.

Para auxiliar na escolha do número de variáveis latentes adequado para o modelo de classificação, foi empregado o procedimento de validação cruzada *leave one out*. Como demonstrado na FIGURA 21a, o aumento no número de variáveis latentes (LVs) só atinge um valor mínimo do erro de previsão (RMSECV) com oito LVs. Contudo, ao se observar o gráfico de variância capturada na matriz X de dados (FIGURA 21b) em função do número de LVs, nota-se que não há um grande aumento no aproveitamento das informações presentes nessa matriz (espectros) acima de três LVs. Portanto, o modelo foi construído utilizando-se três variáveis latentes, as quais explicam 99,15 % da informação contida na matriz X e 92,15 % da informação contida na matriz Y.

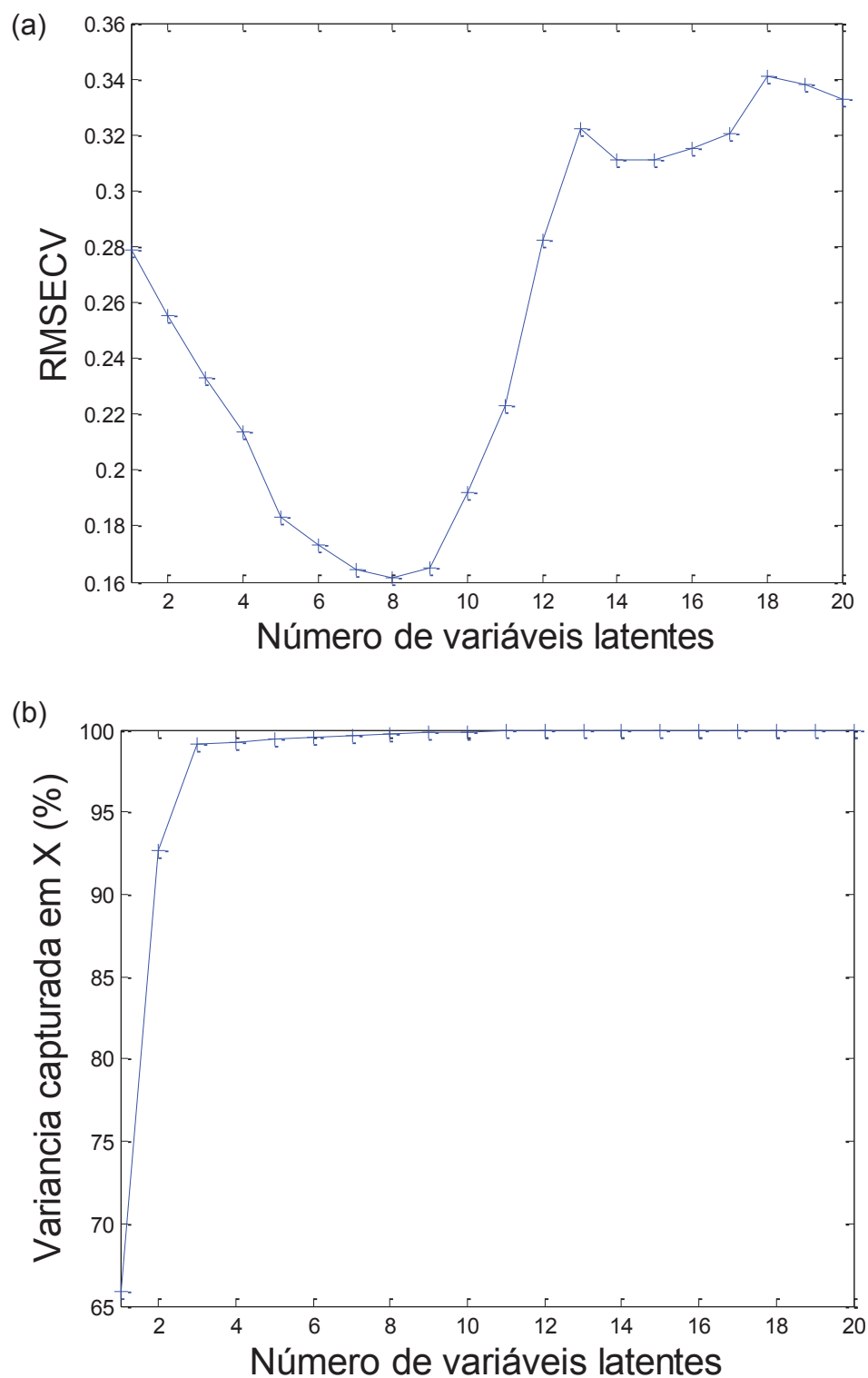


FIGURA 21 – RMSECV VS. NÚMERO DE VARIÁVEIS LATENTES (a). VARIÂNCIA CAPTURADA POR CADA VARIÁVEL LATENTE (b).

A FIGURA 22 representa o vetor de coeficientes de regressão do modelo para três variáveis latentes.

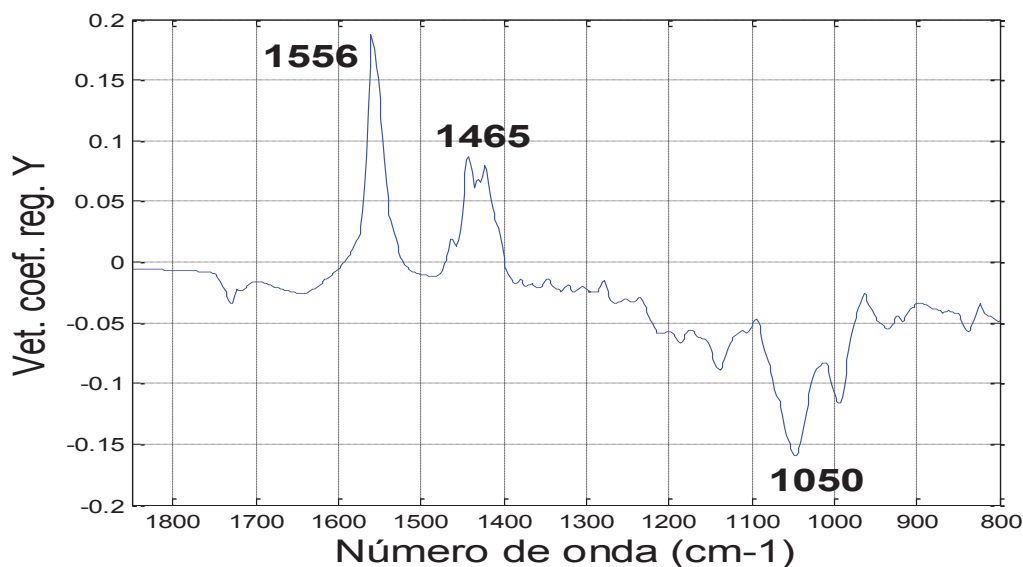


FIGURA 22 – VETOR DE COEFICIENTES DE REGRESSÃO DO MODELO DE PLS-DA PARA CLASSIFICAÇÃO DA ORIGEM DA MATÉRIA PRIMA EMPREGADA EM SABONETES EM BARRA

Um ponto a ser destacado é o fato do vetor de coeficientes de regressão evidenciar as regiões de maior peso na construção do modelo: 1556 cm^{-1} (variações na quantidade de grupos carboxílicos ionizados – sabão), 1465 cm^{-1} (variações nas cadeias carbônicas) e 1050 cm^{-1} (variação no teor de glicerina).

Para detecção de amostras anômalas, foi realizada uma análise de resíduos studentizados *versus leverage*, a qual fornece informações relacionadas à distribuição dos resíduos e o peso de cada amostra em relação ao modelo. Para um nível de confiança de 95 %, segue-se um limite de 2,5 para o valor de resíduos studentizados e, para se definir o valor de limite da *leverage*, utiliza-se a fórmula descrita pela Equação 7.

$$Leverage = \frac{3 \cdot (n^{\circ} VLS)}{(n^{\circ} amostras)} = 0,138 \quad (7)$$

Como pode ser observado na FIGURA 23, apenas uma amostra apresentou valores de resíduos acima do limite de 2,5 (referente aos 95 % de confiança definidos para a análise). Uma vez que a amostra sob avaliação pertence à classe sintética (classe que apresenta as maiores diferenças frente às demais) e que dentro de um nível de confiança de 95 %, assume-se que até 5 % das amostras possam estar fora do limite estabelecido, essas divergências foram consideradas

aceitáveis dentro do escopo do presente trabalho e não foi necessário eliminar nenhuma amostra do conjunto de calibração.

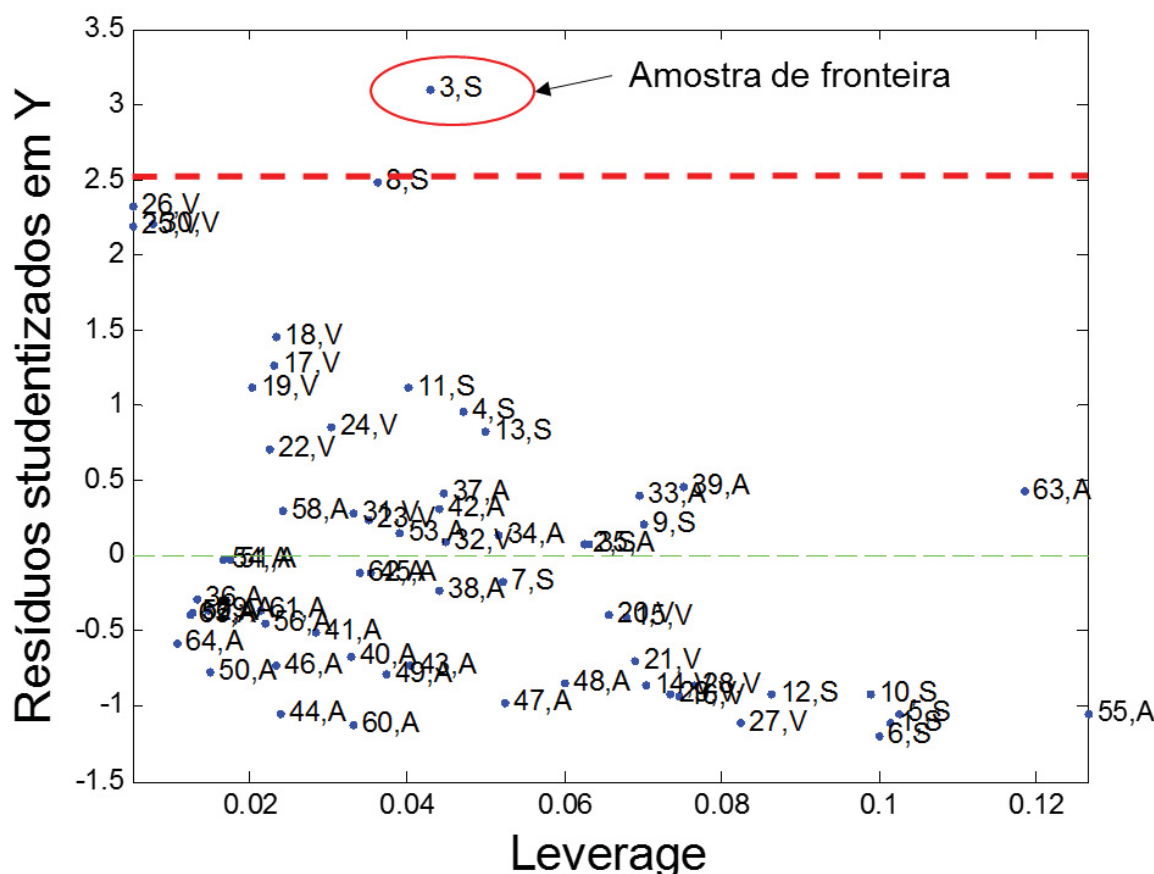


FIGURA 23 – REPRESENTAÇÃO GRÁFICA DA ANÁLISE DE RESÍDUOS STUDENTIZADOS VERSUS LEVERAGE

Em seguida, aplicou-se um método de determinação de fronteiras visando a obtenção de um sistema classificador para as amostras de sabonete. Neste método de classificação, são obtidos limites ou *thresholds* que possibilitam realizar a classificação de amostras que não participaram do conjunto de calibração (conjunto teste).

A FIGURA 24 apresenta os *thresholds* (limites) de 1,60 e 2,66 e a classificação realizada para as amostras do conjunto de calibração via validação cruzada (*leave one out*). Ou seja, amostras que apresentem valor de previsão menor que 1,60 são classificadas como sabonetes que empregaram matéria prima de origem sintética. Amostras que apresentem valor de previsão maior que 2,66 são classificadas como sabonetes contendo matéria prima de origem animal. E

Capítulo IV – Sabonete em barra

amostras que apresentem valores de previsão entre 1,60 e 2,66 são classificadas como sabonetes contendo apenas matéria prima de origem vegetal.

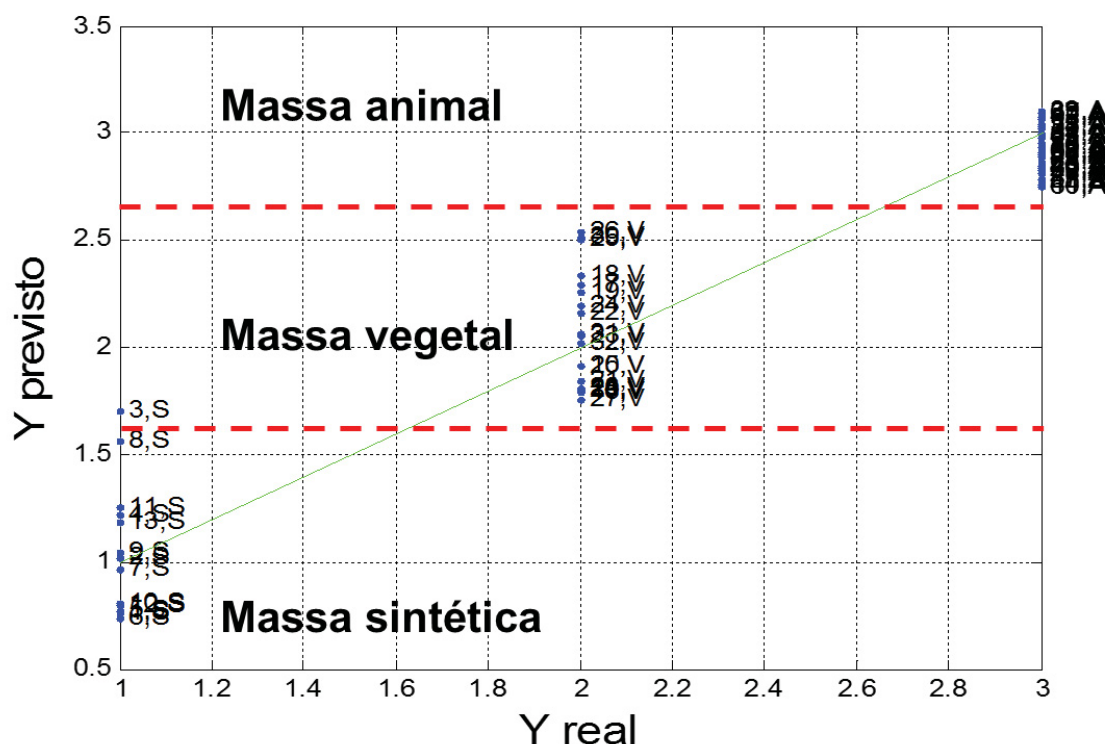


FIGURA 24 - COMPARAÇÃO ENTRE AS CLASSIFICAÇÕES REAIS (DEFINIDAS PELO ESPECIALISTA NO GRUPO BOTICÁRIO) COM AS PREVISTAS PELO MODELO)

Para o conjunto de calibração (validação cruzada), o modelo PLS-DA apresentou um índice de acerto de 98,46 % na diferenciação das amostras de sabonete contendo massa animal, vegetal e sintética (apenas uma amostra sintética foi erroneamente classificada como sendo vegetal).

O modelo construído para o conjunto de calibração foi então empregado para classificar as amostras do conjunto teste (9 amostras cujos espectros não foram empregados na construção do modelo).

A TABELA 13 apresenta os resultados para a classificação do conjunto teste utilizando os *thresholds* definidos para o conjunto de calibração (1,60 e 2,66).

TABELA 13 - VALORES MEDIDOS E PREVISTOS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA

Valores previstos	Classe real	Classe prevista
0,83	Sintética	Sintética
0,97	Sintética	Sintética
1,10	Sintética	Sintética
1,86	Vegetal	Vegetal
2,09	Vegetal	Vegetal
1,92	Vegetal	Vegetal
2,88	Animal	Animal
2,87	Animal	Animal
2,94	Animal	Animal

Os resultados obtidos na classificação para o conjunto teste demonstraram que o modelo construído foi capaz de classificar corretamente todas as 9 amostras selecionadas aleatoriamente dentro de cada grupo. Em outras palavras, comprovou-se a capacidade do método proposto em classificar amostras de sabonetes de acordo com a composição da massa base empregada na produção de sabonetes em barra, mesmo que tais amostras não sejam utilizadas para construção do modelo de calibração.

Entretanto, apesar dos bons resultados, uma limitação do método de classificação por PLS-DA deve ser apontada. O método classifica toda amostra como pertencente a uma classe, de acordo com os valores limites estabelecidos, ou seja, amostras que não pertençam a nenhuma das classes correm o risco de serem classificadas erroneamente. Visando contornar essa limitação, o método SIMCA surge como uma opção interessante, devido ao seu funcionamento baseado na construção de hipercaixas em torno das amostras de cada classe.

7.2.4. SIMCA - Modelagem Independente por Analogia de Classe

Ainda de posse das informações obtidas através da análise de componentes principais e objetivando-se obter a classificação via SIMCA das amostras de sabonetes de interesse nos respectivos grupos de acordo com a composição da matéria prima, foi testado o método SIMCA (do inglês “*Soft Independent Modeling of Class Analogy*”). No SIMCA, cada uma das classes das amostras (no caso,

Capítulo IV – Sabonete em barra

sintético, vegetal e animal) é submetida a uma análise de componentes principais e, após definido o número de componentes principais necessárias para descrever cada classe, é construída uma hipercaixa envolvendo as amostras de cada grupo, cujos limites são definidos com certo intervalo de confiança.

Para a construção do modelo, foram empregados os mesmos pré-processamentos otimizados para a PCA e o PLS-DA, no caso: aproveitamento apenas da região do espectro entre 1850 e 800 cm^{-1} e centrar os dados na média. Para auxiliar na escolha do número de componentes principais, foi empregado o procedimento de validação cruzada *leave one out*. Para a classe 1 (massa sintética), como demonstrado na FIGURA 25a, o aumento no número de componentes principais não leva a grande redução no erro de previsão (PRESS, Soma dos Quadrados dos Erros de Previsão) a partir de quatro PCs. Já na FIGURA 25b, é exibido o gráfico de Q vs T^2 , que representam respectivamente os resíduos e a distância de cada amostra em relação ao centro da hipercaixa para quatro componentes principais para a classe 1 (massa sintética).

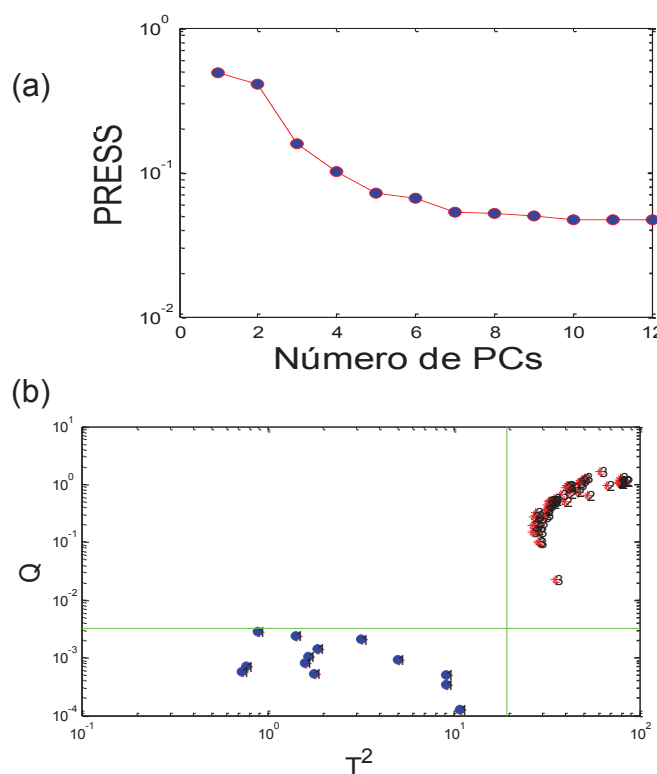


FIGURA 25 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 1 - MASSA SINTÉTICA (a). Q VS T^2 PARA QUATRO COMPONENTES PRINCIPAIS DA CLASSE 1. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 1 E EM VERMELHO AS AMOSTRAS DAS CLASSES 2 E 3 (b)

Capítulo IV – Sabonete em barra

Uma vez que as características dos espectros das amostras pertencentes à classe de base sintética apresentam diferenças mais facilmente detectáveis após a aplicação dos pré-tratamentos, o uso de quatro PCs foi suficiente para a discriminação da classe.

Para a classe 2 (massa vegetal), como demonstrado na FIGURA 26a, o aumento no número de componentes principais não leva a grande redução no erro de predição a partir de quatro PCs. A modelagem empregando quatro componentes principais gerou a distribuição de resíduos e distâncias apresentado na FIGURA 26b.

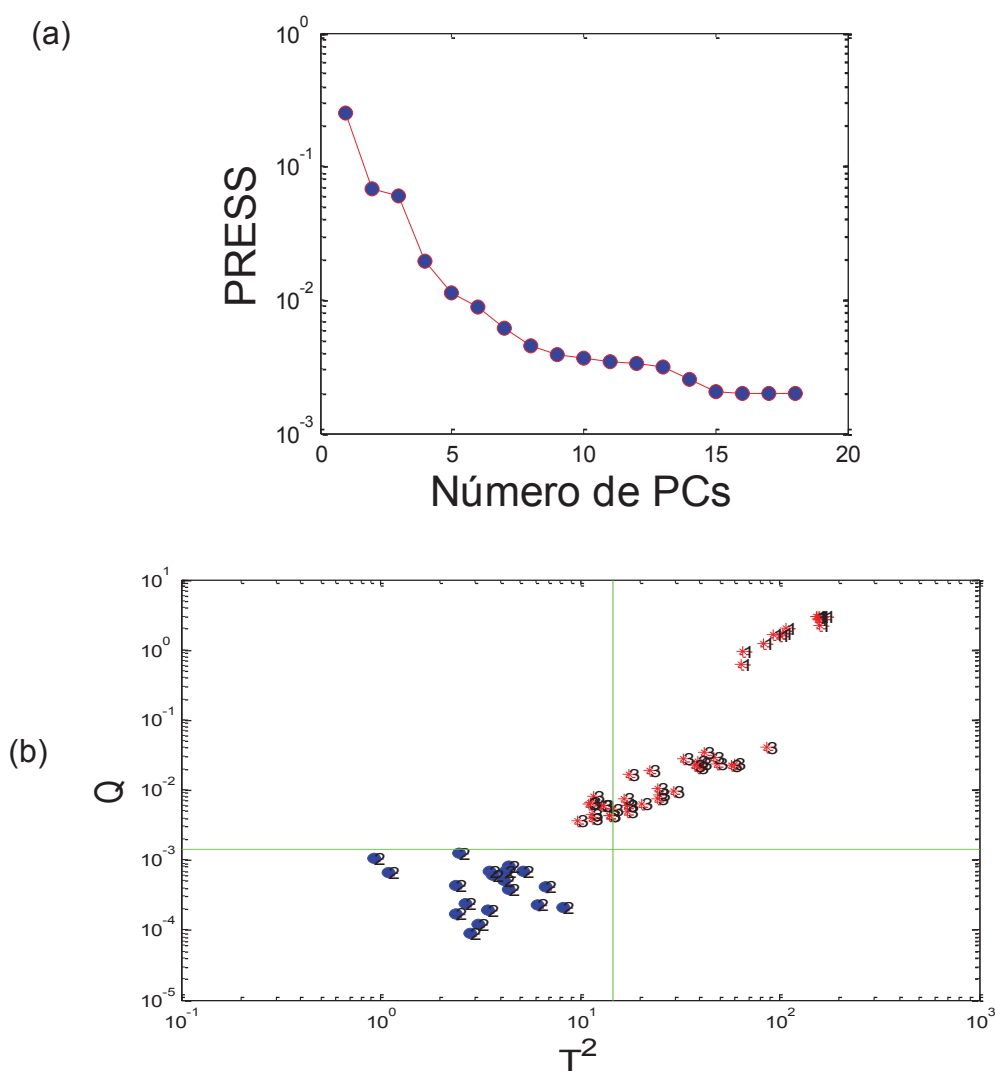


FIGURA 26 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 2 – MASSA VEGETAL (a). Q VS T^2 PARA QUATRO COMPONENTES PRINCIPAIS DA CLASSE 2. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 2 E EM VERMELHO AS AMOSTRAS DAS CLASSES 1 E 3 (b)

Capítulo IV – Sabonete em barra

Mesmo com a elevada semelhança química entre as massas vegetal e animal, e levando-se em consideração que os sabonetes com massa base animal apresentam também massa base vegetal, as amostras da classe vegetal foram separadas das classes animal e sintética com quatro PCs.

Para a classe 3 (massa animal), como demonstrado na FIGURA 27a, o aumento no número de componentes principais não leva a grande redução no erro de predição a partir de três PCs. A modelagem empregando três componentes principais gerou a distribuição de resíduos e distâncias apresentado na FIGURA 27b.

Três componentes principais foi estabelecido como o valor mais adequado para o número de PCs da hipercaixa referente à classe 3 (animal), de maneira que apenas uma amostra apresentou valor de T^2 acima do valor limite.

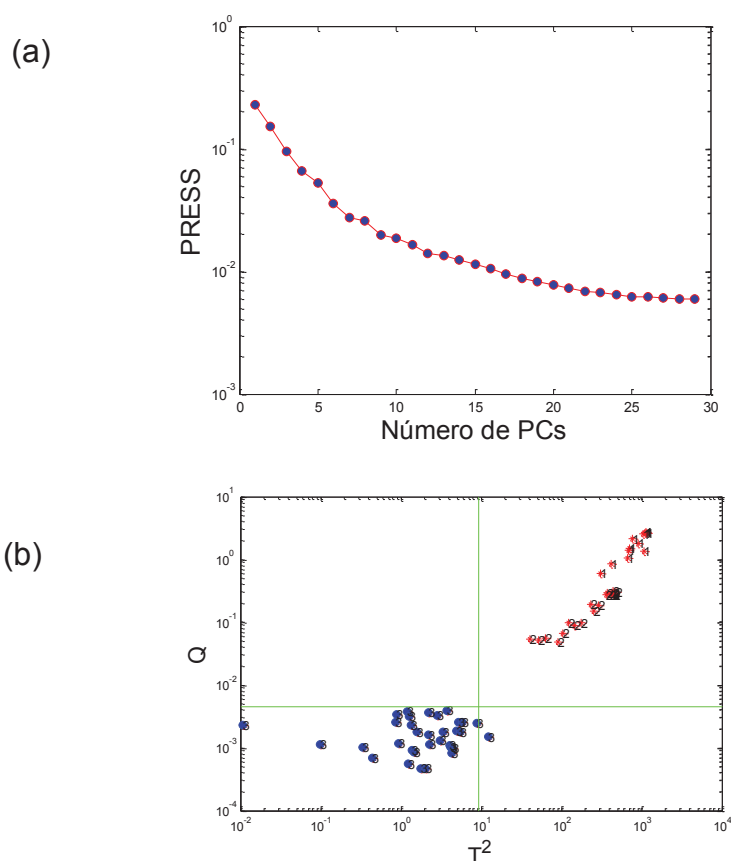


FIGURA 27 - PRESS VS NÚMERO DE COMPONENTES PRINCIPAIS PARA A CLASSE 3 – MASSA ANIMAL (a). Q VS T^2 PARA TRÊS COMPONENTES PRINCIPAIS DA CLASSE 3. EM AZUL AS AMOSTRAS PERTENCENTES À CLASSE 3 E EM VERMELHO AS AMOSTRAS DAS CLASSES 1 E 2 (b)

O passo seguinte consistiu na aplicação do modelo de SIMCA visando à obtenção de um sistema classificador para as amostras de sabonete. Neste modelo

Capítulo IV – Sabonete em barra

de classificação, as hipercaixas construídas com limites multidimensionais definidos, possibilitam realizar a classificação de amostras que não participaram do conjunto de calibração (conjunto teste ou amostras reais). Os resultados do teste de validação interna dentro do conjunto de calibração são apresentados na TABELA 14.

TABELA 14 - CLASSES REAIS E PREVISTAS PARA O CONJUNTO DE VALIDAÇÃO INTERNA

Classe real / Classe prevista	Sintética	Vegetal	Animal
Sintética	13		
Vegetal		19	
Animal			33

Para o conjunto de calibração (validação cruzada), o modelo SIMCA apresentou um índice de acerto de 100 % na diferenciação das amostras de sabonetes.

As informações obtidas pela modelagem via conjunto de calibração com 65 amostras embasaram a construção do modelo para classificar um conjunto teste de amostras não empregadas na construção do mesmo. Esse conjunto foi composto por nove amostras (três de cada classe). Os resultados do teste são apresentados na TABELA 15.

TABELA 15 - CLASSES REAIS E PREVISTAS PARA O CONJUNTO DE VALIDAÇÃO EXTERNA

Classe real/ Classe prevista	Sintética	Vegetal	Animal
Sintética	3		
Vegetal		3	
Animal			3

Os resultados obtidos para o conjunto teste demonstraram que o modelo construído foi capaz de classificar corretamente todas as nove amostras selecionadas aleatoriamente dentro de cada grupo. Adicionalmente, o funcionamento do modelo SIMCA possibilita a detecção de amostras que não se enquadram em nenhum dos grupos, assim como amostras que apresentam características de mais de um grupo. Em outras palavras, comprovou-se a capacidade do método proposto em classificar amostras de sabonetes de acordo com a constituição da massa base de sabonetes em barra. Adicionalmente, foram reforçadas as impressões iniciais sobre a similaridade química das massas animal e vegetal.

7.3. CONSIDERAÇÕES FINAIS E CONCLUSÕES

Os resultados demonstraram a eficiência da utilização conjunta dos métodos quimiométricos – no caso – PCA, PLS-DA e SIMCA na decomposição e calibração de dados obtidos pela técnica de espectroscopia de infravermelho por reflectância total atenuada (ATR) para identificação da constituição da massa base empregada na confecção de sabonetes em barra comerciais.

Ambos os sistemas de classificação propostos (PLS-DA e SIMCA) levaram a resultados de 100 % de acerto na classificação das amostras do conjunto teste analisadas quanto à classe de origem da matéria prima (animal, vegetal ou sintética). O método de PLS-DA gera valores limites de *previsão* que separam os grupos, tornando simples a classificação de novas amostras, enquanto que o SIMCA apresenta uma vantagem estratégica em relação ao método anterior, que seria a possibilidade de classificar amostras como pertencentes a nenhum dos agrupamentos, o que potencialmente facilita a detecção de novas classes em potencial e anomalias nas amostras.

O trabalho proposto apresenta importância analítica, não apenas por representar uma nova metodologia alternativa aos métodos cromatográficos, mas também, por contornar os problemas de interferência de matriz, além de ser não destrutivo, de baixo custo e não envolver a utilização de solventes nem etapas de extração.

8. CONCLUSÕES GERAIS

O presente trabalho explorou uma série de aplicações de ferramentas quimiométricas para a resolução de problemas da indústria cosmética. Foram abordados desafios como a classificação de sabonetes em barra de acordo com a origem da matéria prima (presença de massa base animal, vegetal ou sintética), obtendo-se 100 % de acerto para o conjunto teste contendo amostras comerciais de sabonete em barra que não participaram da etapa de calibração. Adicionalmente, foi desenvolvida uma metodologia de triagem de testes microbiológicos visando à redução nos tempos de desenvolvimento de produtos cosméticos. A metodologia reduziu de 28 para 2 dias o período de comparação de diferentes conservantes para uso em produtos. Foi desenvolvida também uma metodologia quantitativa via PLS, a qual foi bem-sucedida na determinação dos valores de concentração para o conservante Caprilato de Poliglicerila em uma matriz cosmética referente à composição básica de um creme hidratante. Por fim, foi elaborado um sistema baseado nos métodos de HCA e PCA, o qual foi empregado para agrupar 38 emolientes de acordo com as características físico-químicas mais relevantes, sendo capaz de recomendar os melhores substitutos para cada amostra testada.

Embora diversos tenham sido os focos, as frentes de trabalho desenvolvidas demonstraram com sucesso a potencialidade do ferramental quimiométrico na otimização dos resultados obtidos através de técnicas tradicionalmente empregadas na indústria cosmética, reduzindo tempos de desenvolvimento e aumentando o conhecimento e capacidade exploratória dos formuladores.

REFERÊNCIAS

ABIHPEC. Panorama do setor cosmético no Brasil. Disponível em: <<http://www.abihpec.org.br/wp-content/uploads/2014/01/Panorama-do-setor-PORT-10jan2014.pdf>>. Acesso em: 10/4/2014.

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Guia de controle de qualidade de produtos cosméticos. **Revista – Brasília: Anvisa**, p. 19, 2008.

ALMEIDA, M. R.; FIDELIS, C. H. V; BARATA, L. E. S.; POPPI, R. J. Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. **Talanta**, v. 117, p. 305–11, 2013.

BARICELLI, J.; ROCAFULL, M. A.; VÁZQUEZ, D.; et al. β -defensin-2 in breast milk displays a broad antimicrobial activity against pathogenic bacteria. **Jornal de Pediatria (Versão em Português)**, v. 91, n. 1, p. 36–43, 2015.

BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K. Supervised pattern recognition in food analysis. **Journal of chromatography. A**, v. 1158, n. 1-2, p. 196–214, 2007.

BRASIL, M. DA S. **RDC 29: Lista de Substâncias de Ação Conservante permitidas para Produtos de Higiene Pessoal, Cosméticos e Perfumes**. Brasília, 2012.

CARRERO, J. A.; GOIENAGA, N.; FDEZ-ORTIZ DE VALLEJUELO, S.; ARANA, G.; MADARIAGA, J. M. Classification of archaeological pieces into their respective stratum by a chemometric model based on the soil concentration of 25 selected elements. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 65, n. 4, p. 279–286, 2010.

CTFA, C. T. AND F. **Determination of Preservation Adequacy of Water-Miscible Cosmetic and Toiletry Formulations**. Washington, 2001.

DAGO MORALES, Á.; CAVADO OSORIO, A.; FERNÁNDEZ FERNÁNDEZ, R.; DENNES, E. L. Desarrollo de un modelo SIMCA para la clasificación de kerosinas mediante el empleo de la espectroscopía infrarroja. **Química Nova**, v. 31, n. 6, p. 1573–1576, 2008.

DECONINCK, E.; BOTHY, J. L.; DESMEDT, B.; COURSELLE, P.; BEER, J. O. DE. Detection of whitening agents in illegal cosmetics using attenuated total reflectance-infrared spectroscopy. **Journal of Pharmaceutical and Biomedical Analysis**, v. 98, p. 178–185, 2014.

ESCANDAR, G. M.; DAMIANI, P. C.; GOICOECHEA, H. C.; OLIVIERI, A. C. A review of multivariate calibration methods applied to biomedical analysis. **Microchemical Journal**, v. 82, n. 1, p. 29–42, 2006.

FERNÁNDEZ DE CÓRDOVA MANENT, B.; GONZÁLEZ ABELLÁN, E. F. Analysis of Cosmetic Products. **Analysis of Cosmetic Products**. p.29–41, 2007. Elsevier.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, n. 5, p. 724–731, 1999.

FITOUSSI, R. TEM Assessments of the Restructuring Effects of an Emollient Cream on the *Stratum Corneum*. **Journal of Cosmetics, Dermatological Sciences and Applications**, v. 01, n. 04, p. 125–132, 2011.

GENDRIN, C.; ROGGO, Y.; COLLET, C. Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review. **Journal of pharmaceutical and biomedical analysis**, v. 48, n. 3, p. 533–53, 2008.

GILBERT, L.; SAVARY, G.; GRISEL, M.; PICARD, C. Predicting sensory texture properties of cosmetic emulsions by physical measurements. **Chemometrics and Intelligent Laboratory Systems**, v. 124, n. 0, p. 21–31, 2013.

GOLDSCHMIDT. TEGO Cosmo P 813. Disponível em: <http://www.quetzalquimica.com/images/DS_TEGO_Cosmo_P_813_e25-07-2007.pdf>. Acesso em: 14/8/2015.

GUÉGUEN, N.; JACOB, C. Lipstick and tipping behavior: When red lipstick enhance waitresses tips. **International Journal of Hospitality Management**, v. 31, n. 4, p. 1333–1335, 2012.

HALL, M.; FRANK, E.; HOLMES, G.; et al. The WEKA Data Mining Software: An Update. **SIGKDD Explorations**, v. 11, n. 1, 2009.

HANGANU, A.; TODAȘCĂ, M.-C.; CHIRA, N.-A.; MAGANU, M.; ROȘCA, S. The compositional characterisation of Romanian grape seed oils using spectroscopic methods. **Food chemistry**, v. 134, n. 4, p. 2453–8, 2012.

HOLLER, F. J.; SKOOG, D. A.; CROUCH, S. R. **Princípios de Análise Instrumental**. 6ª ed. ed. Bookman, 2009.

JAMES, F. T.; JOHN, L. D.; MICHAEL, M.; et al. Non-molten-mix process for making bar comprising acyl isethionate based solids, soap and optional filler. , 2000. Google Patents.

KANIU, M. I.; ANGEYO, K. H. Challenges in rapid soil quality assessment and opportunities presented by multivariate chemometric energy dispersive X-ray fluorescence and scattering spectroscopy. **Geoderma**, v. 241-242, p. 32–40, 2015.

KAVRAKI, L. E. Dimensionality Reduction Methods for Molecular Motion. **cnx.org**, p. 1–20, 2007.

KHANI, R.; GHASEMI, J. B.; SHEMIRANI, F. Simultaneous multicomponent spectrophotometric monitoring of methyl and propyl parabens using multivariate statistical methods after their preconcentration by robust ionic liquid-based dispersive liquid-liquid microextraction. **Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy**, v. 122, p. 295–303, 2014.

KHANMOHAMMADI, M.; GARMARUDI, A. B. Infrared spectroscopy provides a green analytical chemistry tool for direct diagnosis of cancer. **TrAC Trends in Analytical Chemistry**, v. 30, n. 6, p. 864–874, 2011.

KHANMOHAMMADI, M.; GARMARUDI, A. B.; LA GUARDIA, M. DE. Characterization of petroleum-based products by infrared spectroscopy and chemometrics. **TrAC Trends in Analytical Chemistry**, v. 35, p. 135–149, 2012.

KULIGOWSKI, J.; CASCANT, M.; GARRIGUES, S.; LA GUARDIA, M. DE. An infrared spectroscopic tool for process monitoring: sugar contents during the production of a depilatory formulation. **Talanta**, v. 99, p. 660–7, 2012.

L'ORÉAL GROUP. L'Oréal's history: highlights, stages of development - L'Oréal Group. Disponível em: <<http://www.loreal.com/group/history.aspx>>. Acesso em: 17/4/2014.

LÓPEZ-LÓPEZ, M.; ÖZBEK, N.; GARCÍA-RUIZ, C. Confocal Raman spectroscopy to trace lipstick with their smudges on different surfaces. **Talanta**, v. 123, p. 135–139, 2014.

LÓPEZ-SÁNCHEZ, M.; DOMÍNGUEZ-VIDAL, A.; AYORA-CAÑADA, M. J.; MOLINA-DÍAZ, A. Assessment of dentifrice adulteration with diethylene glycol by means of ATR-FTIR spectroscopy and chemometrics. **Analytica chimica acta**, v. 620, n. 1-2, p. 113–9, 2008.

MERCADANTE, R.; ASSUMPÇÃO, L. DE. Massa base para sabonetes - Fabricando sabonetes sólidos. Disponível em: <<http://projetos.unioeste.br/projetos/gerart/apostilas/apostila7.pdf>>. Acesso em: 20/1/2015.

MIRGHANI, M. E. S.; CHE MAN, Y. B.; JINAP, S.; BAHARIN, B. S.; BAKAR, J. FTIR Spectroscopic Determination of Soap in Refined Vegetable Oils. **Journal of the American Oil Chemists' Society**, v. 79, n. 2, p. 111–116, 2002.

MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, v. 21, n. 4, p. 467–469, 1998.

MORGANO, M. A.; QUEIROZ, S. C. DO N. DE; FERREIRA, M. M. C. Aplicação da Análise exploratória na diferenciação de vegetais. **Brazilian Journal of Food Technology**, v. 2, n. 1, p. 73–79, 1999.

NASCIMENTO, J. A.; DIONÍSIO, A. G. G.; NASCIMENTO, E. C. L. DO; FREITAS, S. K. B.; ARAÚJO, M. C. U. DE. Análise screening de vinhos empregando um analisador fluxo-batelada, espectroscopia UV-VIS e quimiometria. **Química Nova**, v. 33, n. 2, p. 351–357, 2010.

NIE, J.-F.; WU, H.-L.; ZHU, S.-H.; et al. Simultaneous determination of 6-methylcoumarin and 7-methoxycoumarin in cosmetics using three-dimensional excitation-emission matrix fluorescence coupled with second-order calibration methods. **Talanta**, v. 75, n. 5, p. 1260–9, 2008.

O'BRIEN, R. D. **Fats and Oils: Formulating and Processing for Applications**. 3rd ed. 2009.

OLIVERI, P.; DOWNEY, G. Multivariate class modeling for the verification of food-authenticity claims. **TrAC Trends in Analytical Chemistry**, v. 35, p. 74–86, 2012.

PINTO, T. J. A.; KANEKO, T. M.; OHARA, M. T. **Controle biológico de qualidade de produtos farmacêuticos correlatos e cosméticos**. 2ª ed. São Paulo: Atheneu Editora, 2003.

PUTTNAM, N. A.; LEE, S.; BAXTER, B. H. Application of Attenuated Total Reflectance IR spectroscopy to toilet articles and household product. **Journal of Cosmetic Science**, v. 16, p. 607–615, 1965.

R CORE TEAM. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing, Vienna, Austria.**, 2015.

RAJALAHTI, T.; KVALHEIM, O. M. Multivariate data analysis in pharmaceuticals: a tutorial review. **International journal of pharmaceuticals**, v. 417, n. 1-2, p. 280–90, 2011.

ROHMAN, ABDUL; CHE MAN, Y. B. Analysis of lard in cream cosmetics formulations using ft-ir spectroscopy and chemometrics. **Middle-East Journal of Scientific Research**, v. 7, n. 5, p. 726–732, 2011.

SAVARY, G.; GRISEL, M.; PICARD, C. Impact of emollients on the spreading properties of cosmetic products: a combined sensory and instrumental characterization. **Colloids and surfaces. B, Biointerfaces**, v. 102, p. 371–8, 2013.

SILVA, Z. A. DA. **Determinação simultânea de pesticidas utilizando métodos eletroanalíticos e quimiométricos**, 2011. Universidade Federal de São Carlos.

SKOOG, D.; HOLLER, F.; NIEMAN, T. **Principios de análisis instrumental**. 2008.

SMOLIŃSKI, A.; STEMPIN, M.; HOWANIEC, N. Determination of rare earth elements in combustion ashes from selected Polish coal mines by wavelength dispersive X-ray fluorescence spectrometry. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 116, p. 63–74, 2016.

TECPAR. Dossiê Técnico - Fabricação de Sabonete Artesanal. Disponível em: <<http://respostatecnica.org.br/dossie-tecnico/downloadsDT/NTY5NQ==>>. Acesso em: 20/1/2015.

TREVISAN, C. A.; MENDA, M. História dos Cosméticos. Disponível em: <<http://crq4.org.br/?p=texto.php&c=historiadoscsmeticosquimicaviva>>. Acesso em: 10/4/2014.

UARROTA, V. G.; MORESCO, R.; COELHO, B.; et al. Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration. **Food Chemistry**, v. 161, p. 67–78, 2014.

USHIKUBO, F. Y.; CUNHA, R. L. Stability mechanisms of liquid water-in-oil emulsions. **Food Hydrocolloids**, v. 34, p. 145–153, 2014.

WOLD, H. Soft modeling: the basic design and some extensions. **Systems Under Indirect Observation**, v. 2, p. 589–591, 1982.

WORTEL, V. A. L.; WIECHERS, J. W. Skin sensory performance of individual personal care ingredients and marketed personal care products. **Food Quality and Preference**, v. 11, n. 1-2, p. 121–127, 2000